



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Físicas

Escuela Profesional de Física

**Funciones ortogonales empíricas y su aplicación a
datos de temperatura superficial del mar**

TESIS

Para optar el Título Profesional de Licenciado en Física

AUTOR

Edward Steve ALBURQUEQUE SALAZAR

ASESOR

Dr. Joel ROJAS ACUÑA

Lima, Perú

2019



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Alburqueque, E. (2019). *Funciones ortogonales empíricas y su aplicación a datos de temperatura superficial del mar*. Tesis para optar el título profesional de Licenciado en Física. Escuela Profesional de Física, Facultad de Ciencias Físicas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

HOJA DE METADATOS COMPLEMENTARIOS

CÓDIGO ORCID DEL AUTOR: <https://orcid.org/0000-0003-3061-3003>

CÓDIGO ORCID DEL ASESOR: <https://orcid.org/0000-0001-9722-7890>

DNI DEL AUTOR: **43628073**

GRUPO DE INVESTIGACIÓN: **LABORATORIO DE TELEDETECCIÓN - LABTEL**

INSTITUCIÓN QUE FINANCIA PARCIAL O TOTALMENTE LA INVESTIGACIÓN:

UBICACIÓN GEOGRÁFICA DONDE SE DESARROLLÓ LA INVESTIGACIÓN: **UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS. Av. German Amezaga 375, Ciudad Universitaria, Lima, Perú. Latitud: -12,0590208°, Longitud : -77,0822407°**

AÑO O RANGO DE AÑOS QUE LA INVESTIGACIÓN ABARCÓ: **2016-2019**



Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

FACULTAD DE CIENCIAS FÍSICAS

ESCUELA PROFESIONAL DE FÍSICA

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL
TÍTULO PROFESIONAL DE LICENCIADO EN FÍSICA**

Siendo las 16:30 horas del miércoles 17 de julio del 2019, en el Salón de Grados de la Facultad de Ciencias Físicas (Auditorio 109), bajo la Presidencia del Mg. César Aguirre Céspedes (Presidente), Mg. César Omar Jiménez Tintaya (Miembro), Dr. Joel Rojas Acuña (Asesor) y, el Dr. Carlos Augusto Antonio Carbonel Huamán (Suplente), se dio inicio a la Sesión Pública de Sustentación de Tesis para la Licenciatura en Física, mediante la Modalidad M1, del Bachiller:

ALBURQUEQUE SALAZAR, EDWARD STEVE

Dando lectura al Resumen del Expediente, el Presidente del Jurado, invitó al Bachiller Alburqueque Salazar, Edward Steve a realizar una exposición de Trabajo de Tesis titulada: "Funciones ortogonales empíricas y su aplicación a datos de temperatura superficial del mar".

Concluida la exposición del candidato y luego de las preguntas de rigor por parte del Jurado, el Presidente, invitó al Bachiller y al público a abandonar momentáneamente la Sala de Sesión, para dar paso a la deliberación y calificación por parte del Jurado.

Al término de la deliberación del Jurado el Mg. César Aguirre Céspedes, invitó al candidato y al público a pasar a la Sala de Sesión, para dar lectura al resultado de la deliberación. Ha obtenido la calificación de:

Max buena
(MENCIÓN)

18
(NÚMERO)

Dieciocho
(LETRAS)

Finalmente, el Presidente del Jurado, propone al consejo de la Facultad que se le declare Licenciado en Física al Bachiller Alburqueque Salazar, Edward Steve. Siendo las 17:40 Horas, se levanta la Sesión.


Mg. César Aguirre Céspedes
PRESIDENTE


Mg. César Omar Jiménez Tintaya
MIEMBRO


Dr. Joel Rojas Acuña
ASESOR


Dr. Carlos Augusto Carbonel Huamán
SUPLENTE

A mis padres, Segundo y Milagro, por el apoyo y motivación constante, por los consejos, y los valores que me inculcaron. A mi hermano John Jairo por los grandes momentos que pasamos juntos.

Agradecimientos

Quiero iniciar esta tesis agradeciendo a mis padres, quienes siempre me han apoyado en el camino de la ciencia que he elegido, su comprensión ha sido fundamental para salir adelante. A mi asesor, Dr. Joel Rojas Acuña, por sus ideas y consejos para el desarrollo de este trabajo de tesis. Al Dr. Jorge Vazquez del Jet Propulsion Laboratory de la NASA por el tiempo que dedicó a enseñarme diversas técnicas de análisis de datos en IDL. A todos los profesores de la facultad de Ciencias Físicas que contribuyeron con mi formación académica, así como a mis amigos y colegas del Área Funcional de Sensoramiento Remoto del Instituto del Mar del Perú quienes de manera directa o indirecta han colaborado en la finalización de este trabajo de tesis.

ACRÓNIMOS

AF	Análisis de Factores
ATSM	Anomalía de la Temperatura de la Superficie del Mar
ACP	Analisis de Componentes Principales
CEOF	Complex Empirical Ortogonal Functions
CP	Central Pacific
DVS	Descomposición en Valores Singulares
EEOF	Extended Empirical Ortogonal Functions
ENSO	El Niño Southern Oscillation
EMI	El Niño Modoki Index
EP	Eastern Pacific
ERSST	Extended reconstructed sea surface temperature
FOE	Funciones Ortogonales Empíricas
HADISST	Hadley Centre Global Sea Ice and Sea Surface Temperature
HDF	Hierarchical Data Format
IDL	Interactive Data Language
IGP	Instituto Geofísico del Perú
IMARPE	Instituto del Mar del Perú
NETCDF	Network Common Data Form
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OISST	Optimum Interpolation Sea Surface Temperature
TSM	Temperatura de la Superficie del Mar
WMO	World Meteorological Organization

RESUMEN

Posiblemente la técnica estadística más utilizada en ciencias océano-atmosféricas sea el Análisis de las Funciones Ortogonales Empíricas (FOE). El objetivo de aplicar el análisis FOE es reducir un conjunto de datos originales que contiene varias variables a un segundo conjunto de datos que contenga un número mucho menor de variables, con la condición de que estas nuevas variables contengan la máxima fracción posible de la información contenida en el primer conjunto de datos. Este segundo conjunto de datos se obtiene luego de determinar los autovalores y autovectores de la matriz covarianza (Wilks, 1995). El método se ha aplicado a un conjunto de datos de Temperatura de la Superficie del Mar (TSM), estimando primero sus anomalías y construyendo la matriz covarianza, que será utilizada en el método DVS (Descomposición en Valores Singulares) para encontrar los patrones espaciales y temporales (componentes principales).

Se han utilizado datos de nivel Level 4, es decir datos sobre los que se ha realizado un procesamiento adicional tal como el llenado de píxeles faltantes mediante técnicas de interpolación o la combinación de datos de diversos sensores. El área de estudio es el Océano Pacífico Ecuatorial (Latitud: 30°N-30°S, Longitud: 140°E-70°O).

El objetivo es identificar los patrones (también denominados modos) más importantes de la variabilidad de la TSM en el Pacífico Ecuatorial, asociar estos modos con fenómenos como El Niño – Oscilación Sur (ENSO), y posteriormente utilizar las primeras componentes principales para encontrar los índices de variabilidad correspondientes al Pacífico Central (ENSO-CP) y al Pacífico Este (ENSO-EP). Finalmente para validar los índices estimados, se los compara con aquellos obtenidos por instituciones nacionales e internacionales.

Palabras claves: Covarianza, Modos, TSM.

ÍNDICE GENERAL

CAPITULO I. Introducción	1
1.1 Antecedentes	1
1.2 Importancia del análisis FOE	2
1.3 Límites del análisis FOE	3
1.4 Área de estudio	4
1.5 Objetivos	5
CAPITULO II. Funciones Ortogonales Empíricas	6
2.1 Planteamiento del Problema	8
2.2 Terminología	9
2.3 Enfoque geométrico	11
2.4 Enfoque de los autovectores	16
2.5 Relaciones de Ortogonalidad	18
2.6 Matriz Covarianza	19
2.7 Media temporal y espacial	19
2.8 Descomposición en Valores Singulares (DVS)	22
2.9 Factor de escala en los autovectores y componentes principales	23
2.10 Reglas para seleccionar FOE's significantes	27
CAPITULO III. Análisis FOE sobre productos sintéticos	37
CAPITULO IV. Datos y Metodología	38
CAPITULO V. Resultados y Discusión	42
5.1 Variabilidad de la TSM en el Pacífico Tropical	42
5.2 Índices de la diversidad de los Eventos El Niño	45
5.3 Patrones de Variabilidad (no estacionales)	47
5.4 Estimación de los Índices correspondientes al Pacífico Central y Este	56
5.4.1 Estimación de los índices E y C	59
5.4.2 Estimación de los índices EP y CP	64
CAPITULO VI. Conclusiones y Recomendaciones	70
Anexo 1: Fundamentos algebraicos de las Componentes Principales	72
Anexo 2: Repaso de algunos conceptos fundamentales	117
Anexo 3: Funciones Ortogonales Empíricas en IDL	134
REFERENCIAS	143

LISTA DE FIGURAS

- Figura 1.1 Ejemplos en los que el análisis FOE falla (ver las flechas rojas, Shlens (2005))
- Figura 2.1 Modelo propuesto por Shlens (2005) para explicar el funcionamiento del Análisis FOE.
- Figura 2.2. Representación de un objeto desde diferentes perspectivas (Husson et al. 2017)
- Figura 2.3. Formulación del problema del ACP como una minimización de la suma de r_i^2 , donde r_i es la distancia más corta del i-ésimo punto del conjunto de datos hacia el eje z_1 (Hsieh, 2009).
- Figura 2.4. Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella. (Hsieh, 2009)
- Figura 2.5. Rotación de los ejes por un ángulo θ en un espacio bidimensional. (Hsieh, 2009)
- Figura 2.6. Representación vectorial de la componente de \mathbf{y} ortogonal a \mathbf{e}_1 .
- Figura 2.7. Geometría de la regla N (Preisendorfer, 1988). Los modos significativos son aquellos que se encuentran por encima de la curva $\sigma_j(95)$. La figura ilustra el caso en donde el # de modos significativos es igual a 2.
- Figura 3.1. Monopolo espacial
- Figura 3.2. Dipolo Norte-Sur
- Figura 3.3. Dipolo Este-Oeste
- Figura 3.4. Tripolo Este-Oeste
- Figura 3.5. Patrón de propagación
- Figura 3.6. Patrones espaciales y temporales del análisis FOE aplicado al campo = $\text{Data3} + 2 * \text{Data2}$
- Figura 3.7. Patrones espaciales y temporales del análisis FOE aplicado al campo = $\text{Data3} + \text{Data2}$
- Figura 3.8. Patrones espaciales y temporales del análisis FOE aplicado al campo = $\text{Data1} + \text{Data2}$
- Figura 3.9. Patrones espaciales y temporales del análisis FOE aplicado al campo = $\text{Data1} + 2 * \text{Data2}$
- Figura 3.10. Patrones espaciales y temporales del análisis FOE aplicado al campo = Data4
- Figura 3.11. Patrones espaciales y temporales del análisis FOE aplicado al campo = $\text{Data1} + \text{Data2} + \text{Data4}$.

Figura 3.12. Patrones espaciales y temporales del análisis FOE aplicado al campo=
Data2+Data3+ Data5

Figura 3.13. Patrones espaciales y temporales del análisis FOE aplicado al campo=
Data1+Data2+Data3+Data4+Data5

Figura 4.1. Metodología utilizada para obtener las FOE.

Figura 5.1. Compuesto de las anomalías de TSM (datos ERSST) de los eventos El Niño del
Pacífico Este

Figura 5.2. Compuesto de las anomalías de TSM (datos ERSST) de los eventos El Niño del
Pacífico Central

Figura 5.3. Estructura zonal de los compuestos de las anomalías de TSM (datos ERSST) de los
eventos El Niño-PE y el Niño-PC promediado sobre los 2°N - 2°S.

Figura 5.4. Regiones de interés en el Pacífico tropical: Niño 1+2 (0°-10° S, 80°-90° W), Niño 3
(5°S -5°N, 170°W -90° W), y Niño 4 (5°S -5°N, 160°E -150° W). Las anomalías
promedio sobre cada uno de estas tres regiones son usadas como índices. Las
anomalías de TSM del Niño 3 y el Niño 4 son comúnmente usadas como índices
para los episodios de El Niño/La Niña. (Figura obtenida del Centro de Predicción
Climática de la NOAA).

Figura 5.5. Anomalías de las Regiones El Niño

Figura 5.6. Patrón espacial y temporal del modo 1 obtenido de los datos HadISST (1950-2010).

Figura 5.7. Correlación entre el modo 1 (PC1) y las anomalías de TSM de las regiones El Niño.

Figura 5.8. PC1 obtenido de los datos ERSST, HadISST y OISST para el periodo 1950-
2010.

Figura 5.9. Mapas de Correlación y Varianza explicada del primer modo.

Figura 5.10. Patrón espacial y temporal del modo 1 obtenido de los datos HadISST (1950-
2010).

Figura 5.11. Correlación entre el modo 1 (PC2) y las anomalías de TSM de las regiones El
Niño.

Figura 5.12. PC2 obtenido de los datos ERSST, HadISST y OISST para el periodo 1950-2010

Figura 5.13. Mapas de Correlación y Varianza explicada del segundo modo

Figura 5.14. Regiones definidas por Ashok et al., 2007 para encontrar el índice EMI

Figura 5.15. Correlación entre el modo2 (PC2) y el índice EMI

Figura 5.16. Serie temporal del índice EMI

Figura 5.17. Porcentaje de varianza explicada y acumulada por cada modo usando datos HadISST

Figura 5.18. Compuesto de los eventos fuertes El Niño obtenido con el criterio descrito en a)

Figura 5.19: Compuesto de los eventos fuertes La Niña obtenido con el criterio descrito en b)

Figura 5.20: Operación eventos fuertes El Niño – eventos fuertes La Niña

Figura 5.21: Operación eventos fuertes El Niño + eventos fuertes La Niña

Figura 5.22: Gráfico de dispersión de las series temporales PC1 y PC2

Figura 5.23: Series temporales de las primeras 2 componentes principales.

Figura 5.24: Compuestos de la evolución temporal de eventos fuertes El Niño y La Niña

Figura 5.25: Ejes principales de variabilidad de la dispersión de datos PC1 vs PC2

Figura 5.26: Gráfica de dispersión de los nuevos índices obtenidos por rotación (índices E y C)

Figura 5.27: Distribución espacial del Patrón E

Figura 5.28: Distribución espacial del Patrón Patrón C

Figura 5.29: Comparación entre el índice E estimado y el índice E calculado por el IGP.

Figura 5.30: Comparación entre el índice C estimado y el índice C calculado por el IGP.

Figura 5.31: Distribución espacial del Patrón EP

Figura 5.32: Distribución espacial del Patrón CP

Figura 5.33: Evolución temporal de los índices EP y CP

Figura 5.34: Histogramas de los índices EP y CP

Figura 5.35: Coeficientes de correlación desfasada entre las anomalías del Pacífico Tropical y el índice EP

Figura 5.36: Coeficientes de correlación desfasada entre las anomalías del Pacífico Tropical y el índice CP

Figura 5.37: Comparación entre el índice EP estimado y el índice EP calculado por Kao y Yu, 2009.

Figura 5.38: Comparación entre el índice CP estimado y el índice CP calculado por Kao y Yu, 2009.

LISTA DE TABLAS

Tabla 2.1. Guía de la terminología de sinónimos asociados a las FOE (Wilks, 1995).

Tabla 4.1. Productos globales de TSM

Tabla 5.1. Años en que ocurrieron los eventos El Niño-PE y El Niño-PC (Yeh et al., 2009)

CAPÍTULO I

Introducción

Los datos satelitales de TSM obtenidos de agencias internacionales tales como la NASA, NOAA, etc, se caracterizan en general por estar más densamente distribuidos en el espacio que en el tiempo. Por tanto, el análisis para la interpretación de los mismos requiere de técnicas estadísticas que permitan analizar tanto las variaciones temporales como espaciales presentes en los datos. Una de las técnicas para la compresión de la variabilidad de los datos bastante usada en oceanografía es el análisis de Funciones Ortogonales Empíricas (FOE o EOF por sus siglas en ingles). El análisis FOE o también llamado Análisis de componentes principales (ACP) es una de las muchas técnicas inversas y es parecido a un método de reducción de datos ampliamente usado en las ciencias sociales llamado el Análisis de Factores (AF).

1.1 Antecedentes

Tal como lo discute Preisendorfer (1988) en su texto, uno de los aspectos esenciales del método fue desarrollado por un geómetra italiano Beltrami(1873), quien formuló una forma moderna de la resolución de una matriz cuadrada general en su descomposición de valores singulares (DVS). Este mismo descubrimiento fue hecho independientemente por el algebrista francés Jordan (1874). El análisis FOE parece haber hecho su primera aparición en occidente como un ejercicio en álgebra abstracta cuando Sylvester (1889) consideró el problema de la reducción de una matriz cuadrada en sus valores singulares. Una década más tarde Pearson (1901) rediseña el análisis de regresión lineal, en el que introdujo una clara visualización geométrica del método en el espacio euclidiano. La primera aplicación del método en meteorología parece haber sido hecha el Massachusetts Institute of Technology (MIT) por G.P. Wadsworth y sus colegas en 1948. Casi al mismo tiempo un uso completamente independiente del método estaba siendo llevado a cabo por Fukuoka (1951). Cuando la computadora Whirlwind llegó a estar disponible en el MIT en los 50's, E.N. Lorenz a partir del trabajo de Wadsworth y colegas llevó a cabo estudios de predicción de anomalías para

una grilla de 64 puntos cubriendo la parte continental de Estados Unidos, el sur de Canadá y porciones de los océanos circundantes. El trabajo de Lorenz (1956) es ahora un clásico en el campo de la estadística para la predicción climática. Las aplicaciones de las FOE a conjuntos de datos oceanográficos aparecieron casi una década después con el trabajo de Trenberth (1975) en el que relacionó oscilaciones atmosféricas en el Hemisferio Sur con observaciones de TSM. Los estudios de las FOE aplicados a los campos de TSM han sido realizados por diversos autores (Weare, 1976; Lagerloef *et al.*, 1988; Gallaudet *et al.*, 1994; Monahan y Dai 2004; Ashok *et al.* 2007; Takahashi *et al.* 201; Dommenges *et al.* 2013; etc).

1.2 Importancia del análisis FOE

El análisis FOE es importante en el sentido de que datos geofísicos (por ejemplo la TSM) altamente correlacionados pueden ser adecuadamente representados por el menor número de funciones ortogonales y sus coeficientes ortogonales temporales correspondientes (componentes principales). Cuanto mayor es la correlación entre el conjunto de datos, menor es el número de funciones y coeficientes necesarios para describir y explicar la varianza (Davis, 1976; Weare, 1976).

Una propiedad importante de las FOE es que al contrario de otras representaciones ortogonales, ellas no requieren de ninguna forma predeterminada. Una vez que estas funciones son derivadas como autovectores de la matriz covarianza entre las variables, sus formas dependen directamente de las interrelaciones existentes entre los conjuntos de datos analizados. Esta propiedad es especialmente importante cuando se está analizando datos de TSM que no poseen ninguna forma analítica conocida y están sujetos a condiciones de contorno complejas. También, los autovectores (conocidos en aplicaciones geofísicas como los patrones espaciales) obtenidos constituyen una herramienta útil en la interpretación física de los procesos complejos existentes en un conjunto de datos, y sus contribuciones a la varianza total son dadas por los autovalores correspondientes (Kidson, 1975).

Las FOE son ordenadas en orden decreciente, siendo la primer FOE la que contiene la mayor fracción de variancia total. Las FOE de orden mayor muestran pequeñas escalas espaciales con coeficientes caracterizados por menores amplitudes y mayores frecuencias (a veces asociados a ruidos) (Murray *et al.*, 1984). Actualmente en

Meteorología y Oceanografía el análisis FOE ha encontrado una amplia aplicación tanto en el dominio del tiempo como en el dominio de la frecuencia. Datos oceanográficos generalmente exhiben o presentan grandes correlaciones, por lo que luego de aplicar este método se obtiene una representación mucho más compacta de sus variaciones.

1.3 Límites del análisis FOE

El análisis FOE convencional puede ser usado para detectar sólo oscilaciones estacionarias. Para estudiar la propagación de los fenómenos ondulatorios se aplica EEOF (Extended EOF) en la que se necesita usar la matriz de covarianzas rezagadas usada por Weare y Nasstrom (1982), o CEOF (Complex EOF, tópico no abordado en esta tesis) en el dominio de las frecuencias (Wallace y Dickinson, 1972; Horel, 1984). En esta tesis la discusión se enfocará en el análisis FOE convencional en el dominio espacio y tiempo.

Con respecto a las distribuciones espaciales se comentó líneas arriba que una propiedad importante en el análisis FOE era que se podía procesar datos sin importar la distribución que tengan (el análisis FOE no es paramétrico); sin embargo ésta es también una de sus debilidades. El análisis FOE se limita a expresar el conjunto de datos originales como una combinación “lineal” de vectores ortonormales. Es precisamente esta ortonormalidad de los vectores lo que limita el análisis, debido a que con ciertas distribuciones de conjuntos de datos no se obtienen buenos resultados.

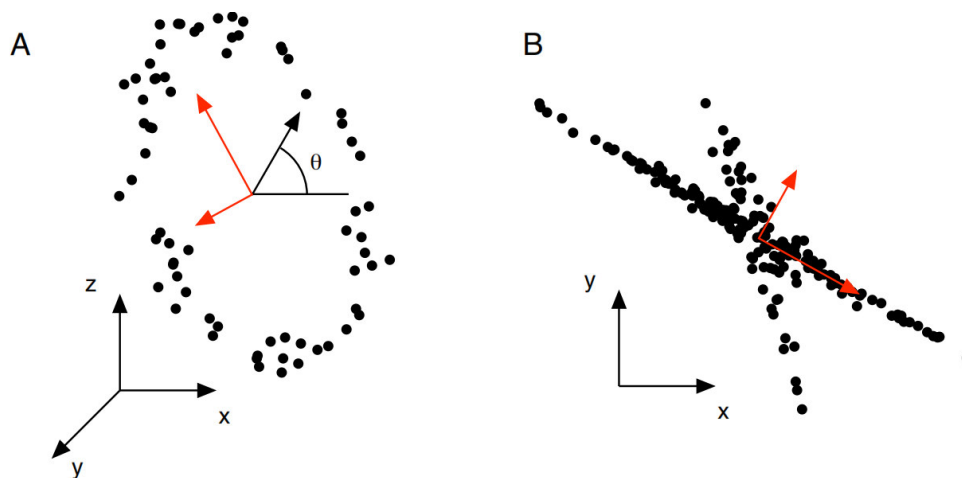


Figura 1.1: Ejemplos en los que el análisis FOE falla (ver las flechas rojas, Shlens (2005))

Shlens (2005) plantea dos ejemplos interesantes. En el primer ejemplo se considera la posición de una persona en una rueda de la fortuna. Si se analizan las posiciones la distribución de los datos es circular, y toda la dinámica puede ser descrita por la variable θ (ángulo de precesión de la rueda). Si el objetivo es encontrar las direcciones principales de variabilidad, será imposible representar la varianza de los datos utilizando componentes ortogonales (Figura 1-A).

El segundo ejemplo consiste en recorrer una ciudad en automóvil: “Conduciendo por la carretera, cuando uno ve otro camino gira hacia la izquierda o la derecha y sigue por ese camino”. En esta analogía el análisis FOE requiere que cada nuevo camino explorado sea perpendicular al anterior, sin embargo este requisito es demasiado estricto ya que comúnmente los caminos en una ciudad están ordenados a lo largo de ejes no ortogonales (Figura 1-B). En este ejemplo el conjunto de datos que tiene distribución no-gausiana y ejes no ortogonales causa que los ejes con la varianza más grande (denominados en el análisis multivariado como “modos” o “patrones”) no correspondan a una respuesta apropiada.

Cuando se aplica el análisis FOE a datos de TSM uno de los principales problemas es la interpretación física de los patrones asociados a la variabilidad de los datos. Estos patrones son contruidos para representar la varianza de una manera óptima (es decir que el mayor porcentaje de varianza de los datos de TSM está contenido en estos patrones), no hay conexiones físicas. Se debe recordar que los patrones son construcciones matemáticas que son escogidos para representar la varianza sobre el dominio de interés, tan eficientemente como sea posible. A menudo las personas esperan obtener “estructuras coherentes” como por ejemplo teleconexiones. También a menudo se piensa que los patrones representan la variabilidad “natural” de un sistema físico. Sin embargo en muchos casos el primer patrón (FOE-1) puede ser identificado como un proceso. Para el segundo patrón (FOE-2) y para los subsiguientes tal asociación es posible solo bajo circunstancias muy especiales (North, 1984).

1.4 Área de Estudio

La TSM es un parámetro relevante para la comprensión y la cuantificación de las interacciones entre el océano y la atmósfera, también es crucial en el desarrollo de los

sistemas biológicos y eventualmente puede usarse como trazador del dinamismo de la superficie. A través de imágenes de satélite se puede estimar este parámetro sinópticamente sobre grandes superficies de manera reiterada. Estas características confieren a este tipo de observaciones un valor incalculable para los estudios oceanográficos. En la zona del Pacífico Ecuatorial la estimación de este parámetro es de mucha importancia debido a que es un indicador de la presencia del fenómeno El Niño, el cual ocurre naturalmente en esta región y que tiene impactos globales de gran relevancia para la sociedad. Además es la región considerada por muchos autores para estimar diversos índices climáticos que ayuden a entender las diferencias entre los eventos tanto en sus patrones espaciales como en sus evoluciones temporales. Por este motivo el área de estudio considerado para aplicar el análisis FOE es el Pacífico Ecuatorial (30° N - 30° S, 140° E - 70° W).

1.5 Objetivos

Los objetivos del trabajo de tesis son:

Objetivo General:

- Reducir la dimensión de los productos de TSM (ERSST, HadISST y OISST) manteniendo la máxima información posible y encontrar los principales patrones de variabilidad del Pacífico Ecuatorial.

Objetivos Específicos:

- Demostrar que con el método FOE se puede asociar los patrones espaciales y temporales de TSM a fenómenos conocidos tales como el Fenómeno El Niño o el Niño Modoki.
- Estimar mediante la rotación de los primeros patrones (ENSO y MODOKI) dos nuevos índices denominados índice E y C los cuales describen la estructura espacial no lineal de los eventos ENSO de una manera óptima.
Estimar mediante la combinación de técnicas EOF y regresión lineal índices que representan los dos tipos de ENSO: Un ENSO del Pacífico Central (ENSO-CP) y otro del Pacífico Este (ENSO-EP).

CAPÍTULO II

Funciones Ortogonales Empíricas

2.1 Planteamiento del Problema

Para entender un determinado fenómeno se requiere de la medición de varias cantidades (espectros, voltajes, velocidades, etc.); desafortunadamente en la vida real uno se enfrenta a problemas relacionados a que las mediciones aparecen con mucho ruido o son redundantes. Este es un obstáculo fundamental en las ciencias empíricas. La motivación para hacer uso del análisis FOE cuando nos enfrentamos con estos problemas se reflejará en el siguiente ejemplo de Física Clásica tomado del artículo de Shlens (2005). Consideremos el movimiento de un resorte ideal. Este sistema consiste en una esfera metálica de masa “ m ” unido a un resorte sin masa y sin fricción (figura 2.1).

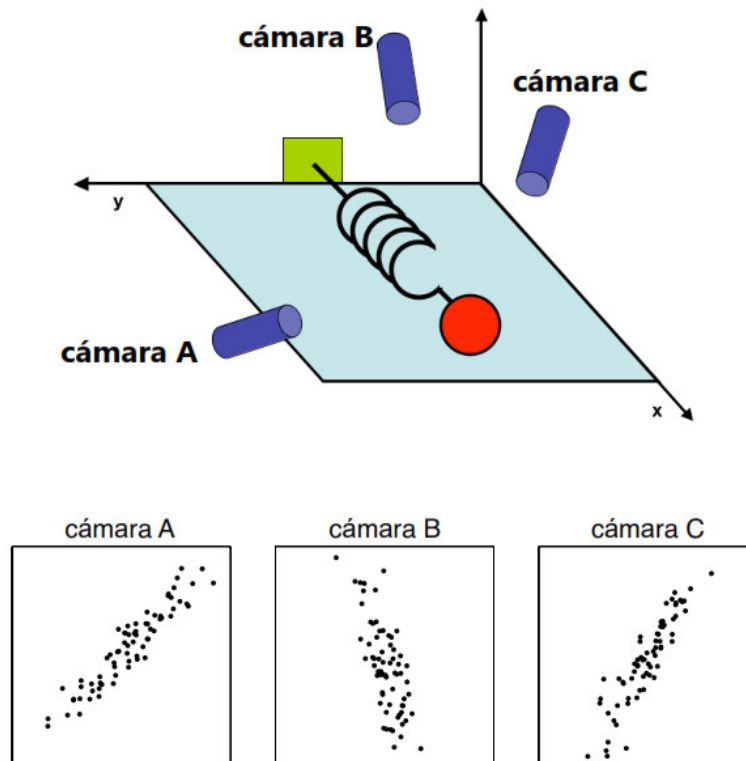


Figura 2.1: Modelo propuesto por Shlens (2005) para explicar el funcionamiento del Análisis FOE. La esfera es alejada una cierta distancia desde el punto de equilibrio (el resorte es estirado). Debido a que el resorte es ideal este oscila indefinidamente a lo largo del

eje x a una frecuencia establecida. Este es un problema en Física básica en el cual el movimiento a lo largo del eje x es resuelto mediante una función que depende del tiempo $f(t) = A\cos(\omega t + \omega_0)$.

Si no tuviéramos ninguna formación en ciencias (experimentadores novatos) no sabríamos nada de esto. No conoceríamos cuántos ejes y dimensiones son importantes para medir. Si nos ponemos en el lugar de un investigador novato decidiríamos medir la posición de la bola en un espacio tridimensional (ya que vivimos en un mundo tridimensional). Específicamente colocaríamos tres cámaras de cine alrededor de nuestro sistema de interés. A una frecuencia determinada cada cámara registra una imagen bidimensional indicando la posición de la esfera (una proyección). Desafortunadamente debido a nuestra ignorancia ni siquiera sabemos cuáles son los ejes ' x ', ' y ' y ' z ', así que elegimos tres posiciones de las cámaras A, B, C con ángulos arbitrarios con respecto al sistema. Los ángulos entre nuestras mediciones ni siquiera son 90° . Lo que sigue es registrar el movimiento de la esfera por varios minutos.

La gran pregunta sería: **¿Cómo conseguimos de estos datos establecer una simple ecuación de x en función del tiempo?**

Conocemos a priori que si fuéramos experimentadores inteligentes habríamos medido la posición a lo largo del eje ' x ' con una sola cámara. Pero esto no es lo que sucede en el mundo real. No sabemos qué medidas reflejan mejor la dinámica de nuestro sistema, registramos más dimensiones de la que realmente necesitamos, además tenemos que lidiar con el problema del ruido (en el problema esto significa que necesitamos lidiar con la fricción del aire, imperfecciones de la cámara o incluso fricción en el resorte nada ideal). El ruido que contamina las mediciones sólo sirve para confundir aún más la dinámica.

Este ejemplo es el desafío que los científicos enfrentan todos los días. Por este motivo el análisis FOE es la herramienta ideal para extraer los patrones espaciales y temporales más importantes asociados a la dinámica de un determinado fenómeno. En este caso aplicaremos el análisis FOE para extraer los principales patrones de variabilidad (tanto temporales como espaciales) de la TSM en la región del Pacífico Ecuatorial.

2.2 Terminología

Este tema de las FOE es considerado como difícil y confuso, siendo el motivo principal de esta confusión la proliferación de la terminología asociada, especialmente en libros o artículos (Wilks, 1995).

La literatura es bastante confusa en lo que respecta a las diferencias entre el método FOE y el método ACP. Aunque algunos autores (Richman, 1986) define los dos métodos diferentes, otros parecen usar el ACP y FOE para describir el mismo método.

• Autovectores(\mathbf{e}_m)	• Autovector	• Componentes Principales(\mathbf{u}_m)	• Componente Principal
• EOFs	• Elementos($\mathbf{e}_{k,m}$)	• Variables Ortogonales Empíricas	• Elementos($\mathbf{u}_{i,m}$)
• Modos de Variación	• Loadings		• Scores($\mathbf{u}_{i,m}$)
• Ejes Principales	• Coeficientes		• Amplitudes
• Vectores Principales	• Pesos Ortogonales Empíricos		• Coeficientes de expansión
• Direcciones Principales			
• Funciones Propias			

Tabla 2.1: Guía de la terminología de sinónimos asociados a las FOE (Wilks, 1995).

Por ejemplo, el método FOE que se describe en el anexo del libro de Peixoto *et al* (1992) es idéntico al método ACP descrita por Preisendorfer (1988). Puesto que la literatura es más o menos confusa sobre la diferencia de estos métodos, se ha optado por utilizar las siglas FOE y ACP indistintamente.

La mayoría de los autores se refieren a los patrones espaciales como FOE, pero algunos se refieren a ellos como las “componentes principales de los patrones”, o simplemente "componentes principales". Las series de tiempo son referidas como “coeficientes de expansión de las series de tiempo”, “coeficientes de expansión”, “componentes principales de las series de tiempo”, o simplemente "componentes principales". En este trabajo de tesis los patrones espaciales serán referidos como FOE y los patrones temporales como las componentes principales. Un resumen de

la diversa terminología que se puede encontrar en la literatura relacionada a las FOE realizada por Wilks (1995) se muestra en la tabla 2.1.

2.3 Enfoque geométrico

Desde un punto de vista geométrico el objetivo del análisis FOE es representar una nube de puntos en un espacio con dimensiones reducidas en una forma “óptima”, es decir “distorsionando las distancias entre individuos lo menos posible”. La figura 2.2 muestra la parte frontal de lo que sería un camello o un dromedario. El punto de vista escogido para esta figura hace que sea un poco complicado identificarlo. *¿Qué es lo que diferencia los puntos de vista entre la parte izquierda y derecha?* En la derecha las distancias están menos distorsionadas y la representación ocupa más espacio en la imagen. La silueta es una proyección de un objeto tridimensional en un espacio bidimensional.

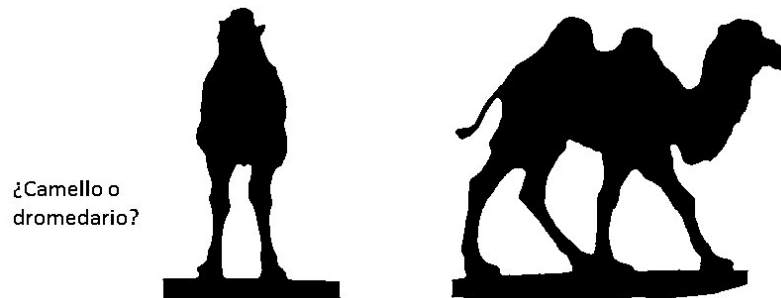


Figura 2.2: Representación de un objeto desde diferentes perspectivas (Husson et al. 2017)

Para que una representación sea satisfactoria se debe seleccionar un punto de vista apropiado. En general FOE significa buscar el mejor espacio de representación (de dimensión reducida), que permite visualizar lo mejor posible la forma de un conjunto de datos.

Usando el ejemplo previo supongamos ahora que se tiene un conjunto de datos con variables (y_1, y_2, \dots, y_m) . Estas variables han sido tomadas ‘n’ veces. En muchas situaciones las ‘m’ variables son ‘m’ series de tiempo que contienen ‘n’ observaciones en el tiempo. Por ejemplo, se puede tener un conjunto de datos que contienen la TSM mensual medida en ‘m’ estaciones durante ‘n’ meses. Si ‘m’ es un número demasiado grande (dimensión original) se podrá capturar la esencia de (y_1, y_2, \dots, y_m) utilizando un nuevo conjunto de variables (z_1, z_2, \dots, z_k) , en

donde $k \ll m$ (k es la dimensión reducida que mejor representa la información contenida en los datos de TSM). Para entender mejor cómo obtener este espacio de dimensión k utilicemos inicialmente un espacio bidimensional. Comencemos con solo dos variables y_1 y y_2 (las variables representan la TSM en diferentes ubicaciones). En la figura 2.3 se muestra el gráfico de dispersión en la que los " n " puntos representan " n " pares de temperaturas. Es evidente que la mayor parte de la varianza en la distribución de los puntos no se encuentra a lo largo de los ejes y_1 y y_2 , pero sí a lo largo del eje z_1 . Si r_i es la distancia entre el i -ésimo punto del conjunto de datos al eje z_1 , entonces el óptimo z_i (proyección del i -ésimo punto al eje z_1) se encuentra al minimizar $\sum_{i=1}^n r_i^2$ (propuesto inicialmente por Pearson en 1901).

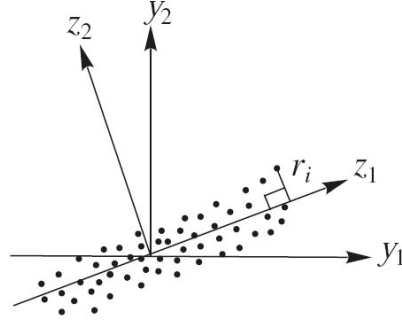


Figura 2.3: Formulación del problema del ACP como una minimización de la suma de r_i^2 , donde r_i es la distancia más corta del i -ésimo punto del conjunto de datos hacia el eje z_1 (Hsieh, 2009).

Para demostrar lo anterior supongamos que se ordenan las dos variables en una matriz \mathbf{X} de dimensiones $n \times 2$, donde las columnas contienen las variables y las filas los elementos. Supondremos que previamente hemos centrado (tipificado) las variables, de manera que cada variable en la matriz \mathbf{X} tiene un valor medio igual a cero y su matriz covarianza está dada por $\mathbf{X}'\mathbf{X}/n$.

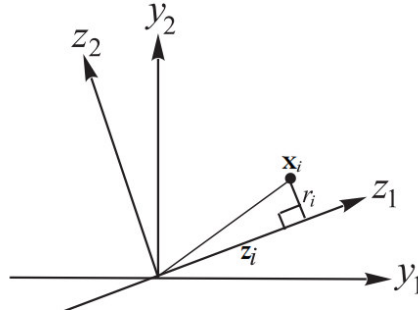


Figura 2.4: Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella. (Hsieh, 2009)

La Figura 2.4 muestra que al proyectar un determinado punto sobre la recta se forma un triángulo rectángulo donde la hipotenusa es la distancia del punto al origen, $(d)^{1/2}$ y los catetos la proyección del punto sobre la recta (z_i) y la distancia entre el punto y su proyección (r_i). Por el teorema de Pitágoras podemos escribir:

$$d = z_i^2 + r_i^2$$

Y sumando esta expresión para todos los puntos

$$\sum_{i=1}^n d = \sum_{i=1}^n z_i^2 + \sum_{i=1}^n r_i^2$$

La condición de que la recta Z_1 capture la mayor parte de la varianza puede concretarse exigiendo que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles.

Como el primer miembro es constante, minimizar $\sum_{i=1}^n r_i^2$ la suma de las distancias a la recta de todos los puntos, es equivalente a maximizar $\sum_{i=1}^n z_i^2$, la suma al cuadrado de los valores de las proyecciones. Como las proyecciones z_i son variables con valor medio cero, maximizar la suma de sus cuadrados equivale a maximizar su varianza, y obtenemos el criterio de encontrar la dirección de proyección Z_1 que maximice la varianza de los datos proyectados.

En tres dimensiones, Z_1 es la mejor línea unidimensional que se ajusta a los puntos, mientras que Z_1 y Z_2 conforman el mejor plano bidimensional que se ajusta a los puntos. En general, con un conjunto de “m” variables se podrá encontrar un ‘hiperplano’ de dimensión ‘k’ ($k \ll m$) que nos dé el mejor ajuste.

2.4 Enfoque de los autovectores

El enfoque más sistemático de los autovectores para el análisis FOE se debe al trabajo realizado por Hotelling (1933). Consideremos nuevamente un ejemplo bidimensional, en el cual un conjunto de datos es transformado desde sus antiguas coordenadas (y_1, y_2) hasta sus nuevas coordenadas (z_1, z_2) mediante una rotación del sistema de coordenadas (Figura 2.5).

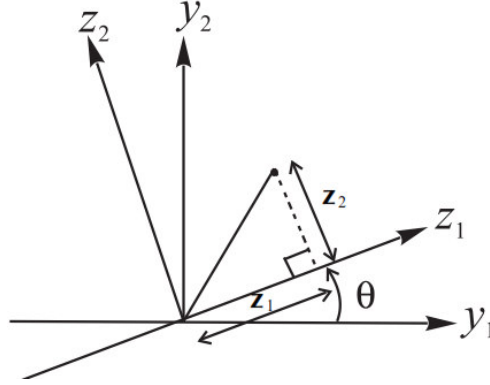


Figura 2.5. Rotación de los ejes por un ángulo θ en un espacio bidimensional (Hsieh, 2009).

$$z_1 = y_1 \cos \theta + y_2 \sin \theta \quad (1)$$

$$z_2 = -y_1 \sin \theta + y_2 \cos \theta$$

En general en un problema de “m” dimensiones, se introducen nuevas coordenadas.

$$z_j = \sum_{l=1}^m e_{jl} y_l, \quad j = 1, \dots, m. \quad (2)$$

El objetivo es hallar

$$\mathbf{e}_1 = [e_{11}, \dots, e_{1m}]^T \quad (3)$$

el cual maximice la varianza de \mathbf{z}_1 , es decir encontrar la transformación de coordenadas de modo que la varianza del conjunto de datos a lo largo de la dirección del eje \mathbf{Z}_1 sea maximizada.

Con

$$\mathbf{z}_1 = \sum_{l=1}^m e_{1l} y_l = \mathbf{e}_1^T \mathbf{y}; \quad \mathbf{y} = [y_1, \dots, y_m]^T, \quad (4)$$

es decir proyectando los elementos de \mathbf{y} en el vector \mathbf{e}_1 obtenemos una distancia de \mathbf{z}_1 a lo largo de la dirección \mathbf{e}_1 . La varianza de \mathbf{z}_1 es:

$$\text{var}(z_1) = \mathbf{E}[(z_1 - \bar{z}_1)(z_1 - \bar{z}_1)] = \mathbf{E}[\mathbf{e}_1^T (\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{e}_1], \quad (5)$$

donde se ha usado la propiedad $\mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$, y en donde \mathbf{E} es el valor esperado definido como:

$$\mathbf{E}(x) = \frac{1}{N-1} \sum_{i=1}^N x_i$$

donde N es el número de medidas de x_i . Continuando con la ecuación (5) tenemos

$$\text{var}(z_1) = \mathbf{e}_1^T \mathbf{E}[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] \mathbf{e}_1 = \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1, \quad (6)$$

donde \mathbf{C} es conocida como la matriz covarianza

$$\mathbf{C} = \mathbf{E}[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] \quad (7)$$

Evidentemente, cuanto más grande sea la norma del vector $\|\mathbf{e}_1\|$, más grande será la varianza de z_1 ($\text{var}(z_1)$). Por lo tanto tenemos que poner una restricción sobre $\|\mathbf{e}_1\|$ mientras intentamos maximizar $\text{var}(z_1)$. Esta restricción de normalización será $\|\mathbf{e}_1\| = 1$, es decir:

$$\mathbf{e}_1^T \mathbf{e}_1 = 1, \quad (8)$$

de esta forma, nuestro problema de optimización consiste en encontrar \mathbf{e}_1 que maximice $\mathbf{e}_1^T \mathbf{C} \mathbf{e}_1$, sujeto a la restricción:

$$\mathbf{e}_1^T \mathbf{e}_1 - 1 = 0 \quad (9)$$

El método de los multiplicadores de Lagrange es comúnmente usado para tratar una optimización sujeta a restricciones. En lugar de encontrar puntos estacionarios de $\mathbf{e}_1^T \mathbf{C} \mathbf{e}_1$, se buscarán los puntos estacionarios de la función de Lagrange L ,

$$L = \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1 - \lambda(\mathbf{e}_1^T \mathbf{e}_1 - 1), \quad (10)$$

donde λ es un multiplicador de Lagrange. Diferenciando L con respecto a \mathbf{e}_1 y luego igualando a cero, obtenemos:

$$\mathbf{C} \mathbf{e}_1 - \lambda \mathbf{e}_1 = 0 \quad (11)$$

Lo cual nos dice que λ y \mathbf{e}_1 son el autovalor y autovector de la matriz de covarianza \mathbf{C} . Multiplicando por \mathbf{e}_1^T a esta ecuación por la izquierda, obtenemos:

$$\begin{aligned} \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1 - \lambda \mathbf{e}_1^T \mathbf{e}_1 &= 0 \\ \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1 - \lambda &= 0 \\ \therefore \lambda &= \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1 = \text{var}(z_1). \end{aligned} \quad (12)$$

Como $\mathbf{e}_1^T \mathbf{C} \mathbf{e}_1$ esta maximizada, también lo están $\text{var}(z_1)$ y λ . La nueva coordenada z_1 , llamada la componente principal, es estimada de la ecuación (4).

El siguiente paso es determinar z_2 , por lo que la tarea será hallar \mathbf{e}_2 que maximice $\text{var}(z_2) = \mathbf{e}_2^T \mathbf{C} \mathbf{e}_2$, sujeto a la restricción $\mathbf{e}_2^T \mathbf{e}_2 = 1$, y a la restricción de que z_2

no debe estar correlacionado con z_1 , es decir la covarianza entre z_2 y z_1 debe ser nula (ecuación 13).

$$\text{cov}(z_1, z_2) = 0. \quad (13)$$

Como una de las propiedades de la matriz covarianza es $\mathbf{C} = \mathbf{C}^T$, podemos escribir:

$$\begin{aligned} 0 &= \text{cov}(z_1, z_2) = \text{cov}(\mathbf{e}_1^T \mathbf{y}, \mathbf{e}_2^T \mathbf{y}) \\ &= \mathbf{E}[\mathbf{e}_1^T (\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{e}_2] = \mathbf{e}_1^T \mathbf{E}[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] \mathbf{e}_2 \\ &= \mathbf{e}_1^T \mathbf{C} \mathbf{e}_2 = \mathbf{e}_2^T \mathbf{C} \mathbf{e}_1 = \mathbf{e}_2^T \lambda_1 \mathbf{e}_1 = \lambda_1 \mathbf{e}_2^T \mathbf{e}_1 = \lambda_1 \mathbf{e}_1^T \mathbf{e}_2 \end{aligned} \quad (14)$$

Por lo tanto la condición de ortogonalidad:

$$\mathbf{e}_2^T \mathbf{e}_1 = 0 \quad (15)$$

puede ser usada como una restricción en lugar de la ecuación (13).

Introduciendo ahora otro multiplicador de Lagrange γ , intentaremos buscar \mathbf{e}_2 , el cual nos dé un punto estacionario de la función de Lagrange L ,

$$L = \mathbf{e}_2^T \mathbf{C} \mathbf{e}_2 - \lambda(\mathbf{e}_2^T \mathbf{e}_2 - 1) - \gamma \mathbf{e}_2^T \mathbf{e}_1. \quad (16)$$

Diferenciando L con respecto a \mathbf{e}_2 y luego igualando a cero, obtenemos:

$$\mathbf{C} \mathbf{e}_2 - \lambda \mathbf{e}_2 - \gamma \mathbf{e}_1 = 0. \quad (17)$$

Multiplicando \mathbf{e}_1^T por la izquierda a la ecuación (17) obtenemos:

$$\mathbf{e}_1^T \mathbf{C} \mathbf{e}_2 - \lambda \mathbf{e}_1^T \mathbf{e}_2 - \gamma \mathbf{e}_1^T \mathbf{e}_1 = 0. \quad (18)$$

Los dos primeros términos de la ecuación (18) son ceros, tomando en cuenta la ecuación (14), mientras que el tercer término es simplemente $\gamma = 0$. Por lo tanto la ecuación (17) se reduce a:

$$\mathbf{C} \mathbf{e}_2 - \lambda \mathbf{e}_2 = 0. \quad (19)$$

Lo cual nos dice que λ y \mathbf{e}_2 son el autovalor y autovector de la matriz covarianza \mathbf{C} . Multiplicando por la izquierda \mathbf{e}_2^T a esta ecuación, obtenemos:

$$\begin{aligned} \mathbf{e}_2^T \mathbf{C} \mathbf{e}_2 - \lambda \mathbf{e}_2^T \mathbf{e}_2 &= 0 \\ \mathbf{e}_2^T \mathbf{C} \mathbf{e}_2 - \lambda &= 0 \\ \therefore \lambda &= \mathbf{e}_2^T \mathbf{C} \mathbf{e}_2 = \text{var}(z_2). \end{aligned} \quad (20)$$

En la ecuación (20) $\lambda = \lambda_2$, con $\lambda_2 < \lambda_1$. Por lo tanto λ_2 es el segundo autovalor más grande (cuando $\lambda_2 = \lambda_1$, se denomina caso degenerado). Todo el proceso que se ha realizado anteriormente puede ser repetido para z_3, z_4, \dots , etc.

¿Cómo podemos conciliar tanto el enfoque geométrico como el enfoque de los autovectores?

Primero sustraemos el valor medio $\bar{\mathbf{y}}$ a \mathbf{y} , de modo que las variables que componen \mathbf{y} estén centradas alrededor del origen con $\bar{\mathbf{y}} = 0$. En el enfoque geométrico minimizamos la distancia entre el conjunto de puntos (cada punto descrito por el vector \mathbf{y}) y los nuevos ejes. Si el vector unitario \mathbf{e}_1 nos da la dirección de los nuevos ejes, entonces la proyección de un punto sobre \mathbf{e}_1 es $(\mathbf{e}_1^T \mathbf{y}) \mathbf{e}_1$. La componente de \mathbf{y} normal a \mathbf{e}_1 es $\mathbf{y} - (\mathbf{e}_1^T \mathbf{y}) \mathbf{e}_1$. Por lo tanto para minimizar la distancia entre los puntos y los nuevos ejes se requiere minimizar la siguiente cantidad:

$$\epsilon = \mathbf{E} [\|\mathbf{y} - (\mathbf{e}_1^T \mathbf{y}) \mathbf{e}_1\|^2] \quad (21)$$

Simplificando:

$$\begin{aligned} \epsilon &= \mathbf{E} [\|\mathbf{y}\|^2 - 2(\mathbf{e}_1^T \mathbf{y})(\mathbf{e}_1^T \mathbf{y}) + (\mathbf{e}_1^T \mathbf{y}) \mathbf{e}_1^T \mathbf{e}_1 (\mathbf{e}_1^T \mathbf{y})] \\ \epsilon &= \mathbf{E} [\|\mathbf{y}\|^2 - (\mathbf{e}_1^T \mathbf{y})^2] = \mathbf{E} [\|\mathbf{y}\|^2] - \mathbf{E} [(\mathbf{e}_1^T \mathbf{y})^2] \\ \epsilon &= \text{var}(\mathbf{y}) - \text{var}(\mathbf{e}_1^T \mathbf{y}), \end{aligned} \quad (22)$$

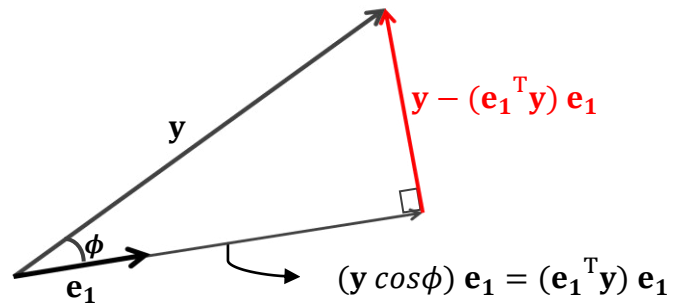


Figura 2.6. Representación vectorial de la componente de \mathbf{y} ortogonal a \mathbf{e}_1 .

donde $\text{var}(\mathbf{y}) = \mathbf{E} [\|\mathbf{y}\|^2]$, asumiendo que $\bar{\mathbf{y}}$ es cero. Como $\text{var}(\mathbf{y})$ es constante, minimizar ϵ es equivalente a maximizar $\text{var}(\mathbf{e}_1^T \mathbf{y})$, lo cual es equivalente a maximizar $\text{var}(z_1)$.

Por lo tanto el enfoque geométrico de minimizar la distancia entre el conjunto de puntos y los nuevos ejes es equivalente al enfoque de los autovectores, encontrando el autovalor más grande λ . Hasta este momento sólo se ha considerado \mathbf{C} como la matriz covarianza; sin embargo esta también puede ser la matriz de correlación si uno lo prefiere. En el análisis de componentes principales combinadas, en donde dos o más variables con diferentes unidades son combinadas en una gran matriz se necesita estandarizar las variables, de modo que \mathbf{C} es la matriz de correlación. Un ejemplo de este caso es cuando se combinan datos de TSM y la presión al nivel del mar para generar la matriz de correlación.

2.5 Relaciones de Ortogonalidad

Con el análisis de componentes principales se encuentran los autovalores y autovectores de la matriz covarianza \mathbf{C} . Los autovectores ortonormales proveen una base, es decir la matriz $\mathbf{y} = [y_1, \dots, y_m]^T$ puede ser expandida en términos de los autovectores $\mathbf{e} = [e_1, \dots, e_j, \dots, e_m]$

$$\mathbf{y} - \bar{\mathbf{y}} = \sum_{j=1}^m a_j(t) \mathbf{e}_j, \quad (22)$$

donde $a_j(t)$ son las componentes principales (patrones temporales). Para obtener $a_j(t)$ multiplicamos \mathbf{e}_i^T por la izquierda en la ecuación anterior, y usamos la relación ortonormal de los autovectores.

$$\mathbf{e}_i^T \mathbf{e}_j = \delta_{ij}, \quad (23)$$

(donde δ_{ij} es la función delta de Kronecker, el cual es 1 si $i = j$ y 0 si $i \neq j$).

Realizada la operación se obtiene:

$$a_j(t) = \mathbf{e}_j^T (\mathbf{y} - \bar{\mathbf{y}}). \quad (24)$$

Es decir que $a_j(t)$ es obtenida por proyección del vector $\mathbf{y} - \bar{\mathbf{y}}$ sobre el autovector \mathbf{e}_j , realizando simplemente un producto escalar entre los dos vectores.

La nomenclatura varía significativamente en la literatura: las a_j son llamadas las componentes principales (CP's o PC's por sus siglas en inglés), coeficientes temporales y amplitudes; mientras que los autovectores \mathbf{e}_j son también referidos

como los vectores principales, patrones espaciales o FOE's, y los j como los modos.

Tener en cuenta que para las series de tiempo a_j es una función del tiempo; mientras que e_j es una función del espacio. Por lo tanto los nombres coeficientes temporales y espaciales están bien descritos. Sin embargo en muchos casos el conjunto de datos puede no consistir de una serie de tiempo. Por ejemplo se puede considerar una imagen satelital multicanal, donde las imágenes de la superficie de la tierra han sido registradas en diferentes longitudes de onda. Aquí t es la ubicación de un pixel en la imagen, y el espacio es indicado por los diferentes canales de longitud de onda (bandas).

Hay dos propiedades muy importantes de las PC's. La expansión $\sum_{j=1}^k a_j(t) e_j(x)$, con $k \leq m$, explica la mayor parte de la varianza contenida en y que cualquier otra combinación lineal $\sum_{j=1}^k b_j(t) f_j(x)$. Por lo tanto ACP provee la forma más eficiente para comprimir cualquier tipo de información en general, usando k autovectores e_j y las correspondientes series de tiempo a_j .

La segunda propiedad importante es que las series de tiempo $\{a_j\}$ no están correlacionadas. Podemos escribir:

$$a_j(t) = e_j^T(y - \bar{y}) = (y - \bar{y})^T e_j, \quad (25)$$

para $i \neq j$,

$$\begin{aligned} \text{cov}(a_i, a_j) &= E[e_i^T(y - \bar{y})(y - \bar{y})^T e_j] = e_i^T E[(y - \bar{y})(y - \bar{y})^T] e_j \\ &= e_i^T C e_j = e_i^T \lambda e_j = \lambda e_i^T e_j = 0. \end{aligned} \quad (26)$$

La ecuación (26) implica una correlación cero entre a_i y a_j , por lo tanto el ACP extrae los patrones no correlacionados de un conjunto de datos. Notar que la no correlación entre a_i y a_j sólo significa que no hay una relación lineal entre los dos; aunque todavía pueden existir relaciones no lineales, las cuales se pueden extraer utilizando el método de ACP no lineal.

Cuando aplicamos ACP a un conjunto de datos raster, uno podría tomar en cuenta la disminución en el área de una celda o grilla con respecto a la latitud.

El área de cada celda es proporcional al coseno de la latitud ϕ . Uno puede evitar que las anomalías en altas latitudes sean sobreestimadas en el ACP. Para ello

multiplicamos la anomalía $\mathbf{y}_l - \bar{\mathbf{y}}_l$ en la l -ésima celda por el factor $(\cos\phi)^{1/2}$ para esa celda (North *et al.*, 1982).

2.6 Matriz covarianza

Consideremos el siguiente conjunto de datos ordenados en la matriz \mathbf{Y}

$$\mathbf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix} \quad (27)$$

donde m es el número de puntos en el espacio y n el número de puntos en el tiempo. Las columnas de la matriz son simplemente los vectores $\mathbf{y}(t_1)$, $\mathbf{y}(t_2)$, $\mathbf{y}(t_n)$. Asumiendo que

$$\frac{1}{n} \sum_{i=1}^n y_{ji} = 0, \quad (28)$$

es decir que la media temporal ha sido removida, entonces

$$\mathbf{C} = \frac{1}{n-1} \mathbf{Y} \mathbf{Y}^T \quad (29)$$

es una matriz $m \times m$. La teoría de la descomposición en valores singulares (DVS) nos dice que los autovalores que no son ceros de $\mathbf{Y} \mathbf{Y}^T$ (una matriz $m \times m$) son exactamente los autovalores que no son ceros de $\mathbf{Y}^T \mathbf{Y}$ (una matriz $n \times n$).

En muchos problemas, el tamaño de las dos matrices es muy diferente. Por ejemplo para datos mensuales de TSM organizados en matrices de dimensión 72×36 (resolución de 5×5 grados) y para un periodo de tiempo de 50 años (600 meses), el número total de datos de temperatura en el espacio es $m=2592$, mientras que el número de datos de temperatura en el tiempo es $n=600$. Obviamente, será mucho más fácil resolver el problema de autovalores para la matriz de tamaño 600×600 que para la matriz de tamaño 2592×2592 .

Por lo tanto cuando $n < m$, es recomendable encontrar en primer lugar los autovalores $\{\lambda_j\}$ y autovectores $\{\mathbf{v}_j\}$ para la matriz de covarianza alternativa

$$\mathbf{C}' = \frac{1}{n-1} \mathbf{Y}^T \mathbf{Y}, \quad (30)$$

es decir

$$\frac{1}{n-1} \mathbf{Y}^T \mathbf{Y} \mathbf{v}_j = \lambda_j \mathbf{v}_j, \quad (31)$$

multiplicando por \mathbf{Y} a la ecuación (31):

$$\mathbf{Y} \left(\frac{1}{n-1} \mathbf{Y}^T \mathbf{Y} \mathbf{v}_j \right) = \mathbf{Y} \lambda_j \mathbf{v}_j = \lambda_j \mathbf{Y} \mathbf{v}_j$$

$$\left(\frac{1}{n-1} \mathbf{Y} \mathbf{Y}^T \right) (\mathbf{Y} \mathbf{v}_j) = \lambda_j (\mathbf{Y} \mathbf{v}_j), \quad (32)$$

esta ecuación puede ser puesta de la siguiente forma

$$\mathbf{C} \mathbf{e}_j = \lambda_j \mathbf{e}_j, \quad (33)$$

con

$$\mathbf{e}_j = \mathbf{Y} \mathbf{v}_j, \quad (34)$$

en donde \mathbf{e}_j es un autovector de \mathbf{C} , y \mathbf{v}_j es un autovector de \mathbf{C}' . En resumen, resolviendo el problema para la matriz más pequeña \mathbf{C}' obtenemos los autovalores $\{\lambda_j\}$ y los autovectores $\{\mathbf{v}_j\}$. Los autovectores $\{\mathbf{e}_j\}$ para la matriz más grande \mathbf{C} son obtenidos de la ecuación (34).

2.7 Media temporal y espacial

Dado un conjunto de datos organizados en la matriz \mathbf{Y} de dimensión $m \times n$ (en donde las filas están conformadas por las variables y las columnas por las muestras obtenidas de cada variable en el tiempo) como en la ecuación (27), se puede remover la media temporal, es decir el valor medio de la j -ésima fila de cada elemento y_{ji} . También se puede remover la media espacial, es decir el valor medio de la i -ésima columna de cada elemento y_{ji} .

El tipo de valor medio que se elija remover depende del tipo de información que se tenga. Para la mayoría de aplicaciones se remueve simplemente la media temporal. Sin embargo para algunas aplicaciones sobre imágenes satelitales es algunas veces más conveniente remover la media espacial a cada uno de los píxeles.

2.8 Descomposición en Valores Singulares (DVS)

En lugar de resolver el problema de autovalores de la matriz de covarianza \mathbf{C} , una manera más conveniente de realizar el ACP desde el punto de vista computacional es aplicar la Descomposición en Valores Singulares (DVS) sobre \mathbf{Y} (definida en la

ecuación 27) de tamaño de $m \times n$ (Kelly, 1988). Sin perder generalidad, podemos asumir que $m \geq n$, siendo la representación del método DVS (Strang, 2005) la siguiente:

$$\mathbf{Y} = \mathbf{E}\mathbf{S}\mathbf{F}^T \quad (35)$$

$$\mathbf{Y} = \mathbf{E}\mathbf{S}\mathbf{F}^T =$$

\mathbf{E}	\mathbf{S}	\mathbf{F}^T
\mathbf{E}' $m \times n$	\mathbf{S}' $n \times n$	
$m \times m$	$m \times n$	

$$\mathbf{Y}_{m \times n} = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix} = \begin{bmatrix} e_{11} & \cdots & e_{m1} \\ e_{12} & \ddots & e_{m2} \\ \vdots & & \vdots \\ e_{1m} & \cdots & e_{mm} \end{bmatrix} \begin{bmatrix} s_{11} & \cdots & 0 \\ 0 & s_{22} & \vdots \\ 0 & \cdots & s_{nn} \\ 0 & & 0 \end{bmatrix} \begin{bmatrix} f_{11} & \cdots & f_{n1} \\ \vdots & \ddots & \vdots \\ f_{1n} & \cdots & f_{nn} \end{bmatrix}$$

\mathbf{E} es una matriz $m \times m$ que contiene una sub-matriz \mathbf{E}' de tamaño $m \times n$ y si $m > n$ algunas columnas de vectores son ceros. \mathbf{S} es una matriz $m \times n$ que contiene una Sub-matriz diagonal \mathbf{S}' de tamaño $n \times n$, y posiblemente algunas filas de vectores son ceros. \mathbf{F}^T es una matriz $n \times n$. Las matrices \mathbf{E} y \mathbf{F} son matrices ortonormales, es decir ellas satisfacen:

$$\mathbf{E}^T \mathbf{E} = \mathbf{I}, \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}, \quad (36)$$

donde \mathbf{I} es la matriz identidad. Las n columnas más a la izquierda de \mathbf{E} contienen los n vectores singulares izquierdos, y las columnas de \mathbf{F} los n vectores singulares derechos, mientras que los elementos diagonales de \mathbf{S}' son los valores singulares.

La matriz covarianza \mathbf{C} puede ser reescrita como

$$\mathbf{C} = \frac{1}{n} \mathbf{Y}\mathbf{Y}^T = \frac{1}{n} (\mathbf{E}\mathbf{S}\mathbf{F}^T) (\mathbf{E}\mathbf{S}\mathbf{F}^T)^T$$

$$\mathbf{C} = \frac{1}{n}(\mathbf{E}\mathbf{S}\mathbf{F}^T)(\mathbf{F}\mathbf{S}^T\mathbf{E}^T) = \frac{1}{n}\mathbf{E}\mathbf{S}\mathbf{S}^T\mathbf{E}^T \quad (37)$$

La matriz

$$\mathbf{S}\mathbf{S}^T = \mathbf{\Lambda} . \quad (38)$$

Es diagonal y cero en todas partes, excepto en la esquina superior izquierda $n \times n$, conteniendo \mathbf{S}'^2 .

Multiplicando por la derecha a la ecuación (37) por $n\mathbf{E}$, y usando las ecuaciones (36) y (38), se obtiene:

$$\begin{aligned} \mathbf{C}n\mathbf{E} &= \frac{1}{n}\mathbf{E}\mathbf{S}\mathbf{S}^T\mathbf{E}^Tn\mathbf{E} \\ n\mathbf{C}\mathbf{E} &= \mathbf{E}\mathbf{S}\mathbf{S}^T\mathbf{E}^T\mathbf{E} = \mathbf{E}\mathbf{S}\mathbf{S}^T \\ n\mathbf{C}\mathbf{E} &= \mathbf{E}\mathbf{\Lambda} \end{aligned} \quad (39)$$

En la ecuación (39) $\mathbf{\Lambda}$ contiene los autovalores de la matriz $n\mathbf{C}$. En vez de resolver el problema (39), usaremos el método DVS para obtener \mathbf{E} de \mathbf{Y} mediante la ecuación (35). La ecuación (39) implica que hay solo n autovalores en $\mathbf{\Lambda}$ a partir de \mathbf{S}'^2 , y los autovalores = (valores singulares)². Como (39) y (33) son equivalentes excepto por la constante n , los autovalores en $\mathbf{\Lambda}$ son simplemente $n\lambda_j$, siendo λ_j los autovalores de (33).

Similarmente la otra matriz de covarianza

$$\mathbf{C}' = \frac{1}{n}\mathbf{Y}^T\mathbf{Y} \quad (40)$$

puede ser reescrita como:

$$\begin{aligned} \mathbf{C}' &= \frac{1}{n}(\mathbf{F}\mathbf{S}^T\mathbf{E}^T)(\mathbf{E}\mathbf{S}\mathbf{F}^T) = \frac{1}{n}\mathbf{F}\mathbf{S}^T\mathbf{S}\mathbf{F}^T \\ \mathbf{C}' &= \frac{1}{n}\mathbf{F}\mathbf{S}'^2\mathbf{F}^T \end{aligned} \quad (41)$$

Y finalmente

$$n\mathbf{C}'\mathbf{F} = \mathbf{F}\mathbf{S}'^2, \quad (42)$$

por lo tanto la ecuación (42) tiene los mismos autovalores que la ecuación (39).

La descomposición ACP

$$\mathbf{y}(t) = \sum_j \mathbf{e}_j a_j(t) \quad (43)$$

es equivalente a la matriz de la forma

$$\mathbf{Y} = \mathbf{E}\mathbf{A}^T = \sum_j \mathbf{e}_j \mathbf{a}_j^T \quad (44)$$

donde el autovector \mathbf{e}_j es la j -ésima columna en la matriz \mathbf{E} , y la PC $\mathbf{a}_j(t)$ es el vector \mathbf{a}_j , la j -ésima columna en la matriz \mathbf{A} . Las ecuaciones (35) y (44) establecen que

$$\mathbf{A}^T = \mathbf{S}\mathbf{F}^T \quad \text{o} \quad \mathbf{A}^T = \mathbf{E}^T\mathbf{Y} \quad (45)$$

Kelly (1988) señala que el método DVS es al menos dos veces más rápido que el cálculo de autovectores y autovalores convencional.

2.9 Factor de escala en los autovectores y componentes principales

Hay varias opciones para el escalamiento de las PC's $\{\mathbf{a}_j(t)\}$ y los autovectores $\{\mathbf{e}_j\}$. Uno puede introducir un factor de escala arbitrario α .

$$\mathbf{a}'_j = \frac{1}{\alpha} \mathbf{a}_j ; \quad \mathbf{e}'_j = \alpha \mathbf{e}_j \quad (46)$$

de modo que:

$$\mathbf{y} - \bar{\mathbf{y}} = \sum_{j=1}^m \mathbf{a}_j(t) \mathbf{e}_j = \sum_{j=1}^m \mathbf{a}'_j(t) \mathbf{e}'_j, \quad (47)$$

con $\alpha = -1$ se invierten los signos tanto de \mathbf{a}_j como de \mathbf{e}_j , que a menudo se realiza para hacerlos más interpretables.

El escalamiento ha sido hasta ahora:

$$\mathbf{e}_i^T \mathbf{e}_j = \delta_{ij} \quad (48)$$

el cual fue la elección realizada por Lorenz (1956). La varianza " \mathbf{y} " está por lo tanto contenida en $\{\mathbf{a}_j(t)\}$, con

$$\text{var}(\mathbf{y}) = \mathbf{E} \left[\sum_{j=1}^m \mathbf{a}_j^2 \right] \quad (49)$$

Otra elección común es la que eligió originalmente Hotelling (1933).

$$\mathbf{a}'_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{a}_j ; \quad \mathbf{e}'_j = \sqrt{\lambda_j} \mathbf{e}_j. \quad (50)$$

Se calcula la varianza de \mathbf{a}'_j para demostrar la importancia de este escalado

$$\begin{aligned}
\text{var}(a'_j) &= \text{var}\left(\frac{1}{\sqrt{\lambda_j}} a_j\right) = \text{var}\left(\frac{\mathbf{e}_j^T \mathbf{y}}{\sqrt{\lambda_j}}\right) \\
&= \left(\frac{\mathbf{e}_j^T \mathbf{y}}{\sqrt{\lambda_j}}\right) \left(\frac{\mathbf{e}_j^T \mathbf{y}}{\sqrt{\lambda_j}}\right)^T = \frac{1}{\lambda_j} \mathbf{e}_j^T \mathbf{y} \mathbf{y}^T \mathbf{e}_j = \frac{1}{\lambda_j} \mathbf{e}_j^T \mathbf{C} \mathbf{e}_j = \frac{1}{\lambda_j} \mathbf{e}_j^T \lambda_j \mathbf{e}_j = 1 \\
&\rightarrow \text{var}(a'_j) = 1,
\end{aligned}$$

con lo cual ahora los autovectores escalados \mathbf{e}'_j llevan las unidades de " \mathbf{y} ", representando un típico patrón de anomalías.

$$\text{var}(\mathbf{y}) = \sum_{j=1}^m \lambda_j = \sum_{j=1}^m \|\mathbf{e}'_j\|^2 \quad (51)$$

$$\text{cov}(a'_i, a'_j) = \delta_{ij} \quad (52)$$

La varianza de " \mathbf{y} " está ahora contenida en $\{\mathbf{e}'_j\}$. En resumen, independientemente del factor de escala que se elija, los autovectores son ortogonales y las PCs no están correlacionadas.

Si el ACP es realizado sobre variables estandarizadas $\tilde{\mathbf{y}}_l$, es decir \mathbf{y}_l con la media removida y normalizada por la desviación estándar, entonces se puede demostrar la correlación

$$\rho(a'_j(t), \tilde{\mathbf{y}}_l(t)) = \mathbf{e}'_{jl}, \quad (53)$$

donde ρ es coeficiente de correlación de Pearson, el cual es definido como:

$$\rho_{xy} = \frac{\text{cov}(x, y)}{S_x S_y},$$

en donde S_x y S_y son las desviaciones estándar para x e y respectivamente.

En la ecuación (53) el l -ésimo elemento de \mathbf{e}'_j (Jolliffe, 2002) convenientemente provee la correlación entre la PC \mathbf{a}'_j y la variable estandarizada $\tilde{\mathbf{y}}_l$, que es la razón por la cual el escalado de Hotelling (ecuación 50) es ampliamente usado.

2.10 Reglas para seleccionar FOE's significantes

En la práctica, los modos con menor varianza asociada, los cuales básicamente contienen ruido, son descartados. ¿De qué manera decidimos qué número de modos hay que retener?. Hay algunas 'reglas de oro' o 'reglas de selección'. Uno de los enfoques más simples pero muy subjetivo es graficar los autovalores λ_j en

función del número de modos j . Lo que se espera es encontrar una transición abrupta entre autovalores grandes y pequeños alrededor del modo k . Si esto ocurre uno puede retener los primeros k modos.

A continuación consideraremos algunos procedimientos más avanzados para separar la señal del ruido.

2.10.1 Regla de Selección N de Preisendorfer

La regla N de Preisendorfer (Overland y Preisendorfer, 1981) está basada en el método de Monte Carlo, que implica la creación de matrices de datos aleatorios $\mathbf{R}_l (l = 1, \dots, L)$, del mismo tamaño que la matriz \mathbf{Y} . Los elementos aleatorios deben estar normalmente distribuidos y deben tener la misma varianza de los datos reales. El análisis FOE es aplicado sobre cada una de las matrices aleatorias, obteniendo los autovalores $\lambda_j^{(l)}$. Asumimos que para cada l , el conjunto de autovalores es ordenado en orden descendiente $\lambda_j^{(l)} = [\lambda_1^{(l)}, \lambda_2^{(l)}, \lambda_3^{(l)} \dots \dots \lambda_p^{(l)}]$. Por lo que tendríamos una matriz de autovalores de la siguiente forma:

l	1	2	p
1	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\lambda_p^{(1)}$
2	$\lambda_1^{(2)}$	$\lambda_2^{(2)}$	$\lambda_p^{(2)}$
\vdots	\vdots	\vdots	\vdots
100	$\lambda_1^{(100)}$	$\lambda_2^{(100)}$	$\lambda_p^{(100)}$

Ahora calculamos el % de varianza explicada de los autovalores como:

$$U_j(l) = \lambda_j^{(l)} \left[\sum_{j=1}^p \lambda_j^{(l)} \right]^{-1} \times 100 \quad (54)$$

Para $l = 1, \dots, L$ y $j = 1, \dots, p$. Reconstruyendo nuestra tabla anterior tenemos:

l	1	2	p
1	$U_1(1)$	$U_2(1)$	$U_p(1)$
2	$U_1(2)$	$U_2(2)$	$U_p(2)$
\vdots	\vdots	\vdots	\vdots
100	$U_1(100)$	$U_2(100)$	$U_p(100)$

Se está asumiendo según la tabla anterior que hemos generado 100 matrices aleatorias ($L = 100$), aunque podrían ser más, dependiendo del tamaño de la matriz de datos original y de la capacidad de nuestra computadora.

Para cada orden j ($j = 1, \dots, p$) se reordenan los L autovalores $U_j(l)$ en forma creciente, de forma tal que

$$U'_1(001) < \dots < U'_1(100)$$

$$U'_2(001) < \dots < U'_2(100)$$

$$\vdots$$

$$U'_p(001) < \dots < U'_p(100)$$

Luego fijamos como nivel de significación el 95% para cada orden j , es decir los puntos correspondientes a $U'_j(095)$.

Sea $\sigma_j(95) = U'_j(005)$, el paso final consiste en graficar $\sigma_j(95)$ y superponer el % de la varianza explicada de los autovalores de la matriz original \mathbf{Y} (figura 2.7), es decir:

$$V_j = \lambda_j \left[\sum_{j=1}^p \lambda_j \right]^{-1} \times 100 \quad (55)$$

El número de modos significativos será igual al # de valores $V_j/\sigma_j(95)$ que sean mayores a 1

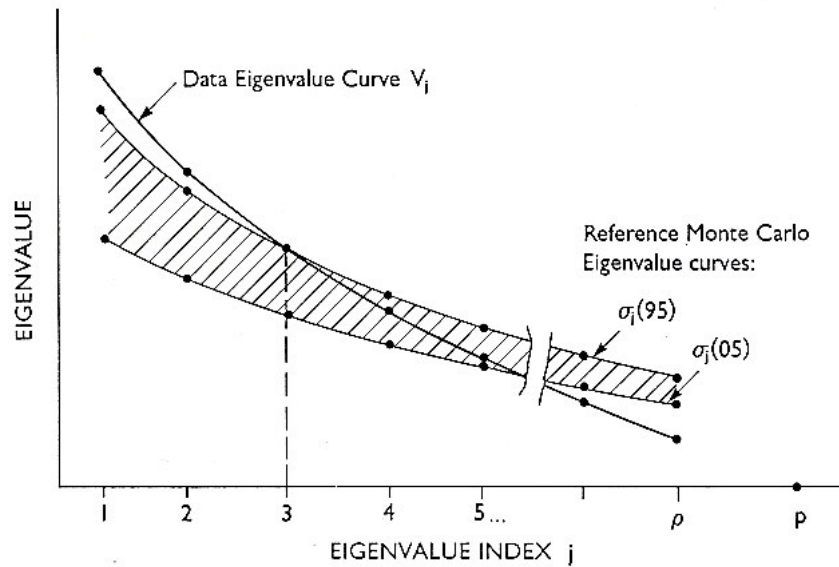


Figura 2.7. Geometría de la regla N (Preisendorfer, 1988). Los modos significativos son aquellos que se encuentran por encima de la curva $\sigma_j(95)$. La figura ilustra el caso en donde el # de modos significativos es igual a 2.

2.10.2 Regla de North

Según North et al. (1982), el error en la estimación del autovalor λ_α , está dado por

$$\delta\lambda_\alpha \approx \sqrt{\frac{2}{N}} \lambda_\alpha \quad (56)$$

donde N es el número de realizaciones. Si otro autovalor λ_β está muy próximo a λ_α tal que:

$$\Delta_{\alpha\beta} = |\lambda_\alpha - \lambda_\beta| \leq \delta\lambda_\alpha \quad (57)$$

Es decir, si la diferencia entre autovalores es comparable al error de estimación, ocurre una “*degeneración efectiva*”. La degeneración de dos autovalores conduce a una ambigüedad intrínseca en la definición de los autovectores y componentes correspondientes, pues cualquier combinación lineal de los autovectores en cuestión, sería también un autovector.

El número de modos significativos será igual a los primeros autovalores que no cumplan con la ecuación (57).

CAPÍTULO III

Análisis FOE sobre productos Sintéticos

Con el fin de entender cómo trabaja el análisis FOE sobre conjuntos de datos, se ha realizado una prueba aplicada a productos sintéticos. Para la generación de estos datos se ha utilizado el lenguaje de programación IDL (Interactive Data Language) y se han reproducido los ejemplos dados en Venegas (2001).

El conjunto de datos (distribuidos en el espacio y tiempo) está conformado por series de tiempo de 1200 muestras sobre una grilla de 41 x 41 puntos (1681 series de tiempo en total). Las 1200 muestras representan 1200 meses (100 años).

El primer producto generado denominado “DATA1” está conformado por un monopolio espacial centrado en el punto medio del dominio, el cual oscila con un periodo de 15 años (180 meses). Su estructura espacial es dada en la figura 3.1 junto con su evolución temporal.

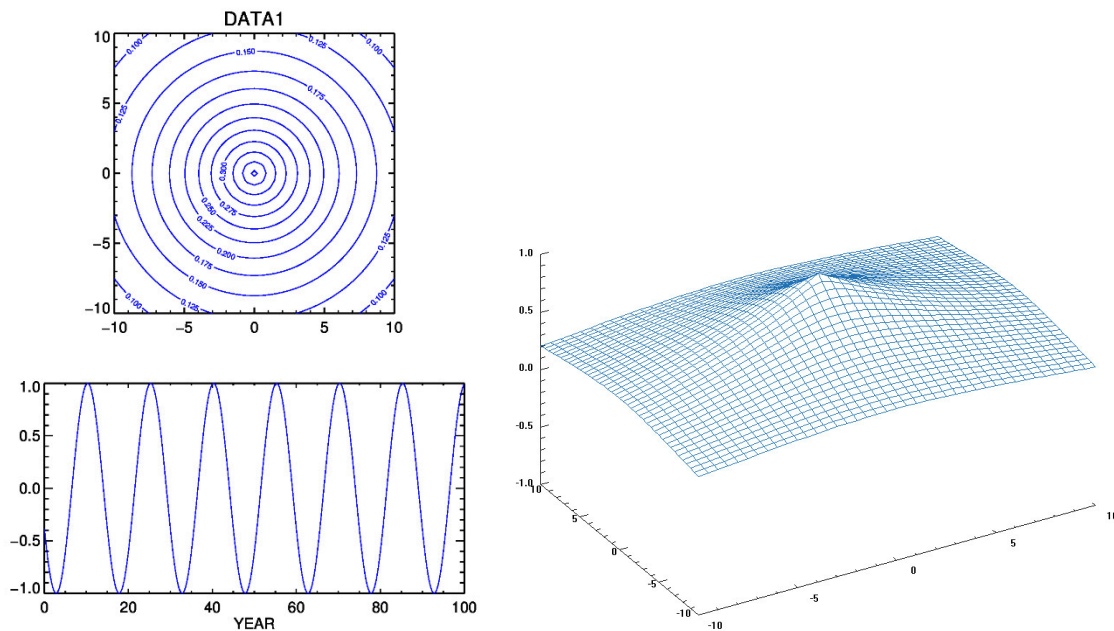


Figura 3.1. Monopolio espacial

El segundo producto “DATA2” tiene un patrón de dipolo norte-sur en el espacio y oscila con un periodo de 5 años (60 meses). La figura 3.2 muestra su estructura espacial y su serie de tiempo asociada.

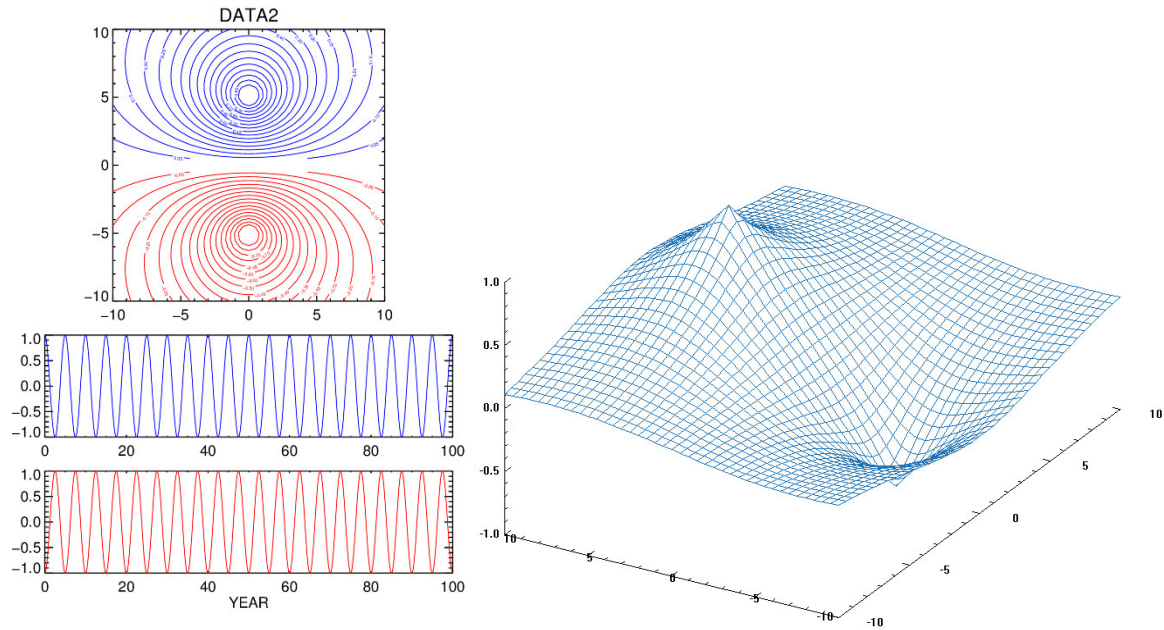


Figura 3.2. Dipolo Norte-Sur

El tercer producto “DATA3” presenta un dipolo este-oeste en el espacio y oscila con un periodo de 15 años (180 meses). La figura 3.3 muestra su estructura espacial y su correspondiente serie de tiempo.

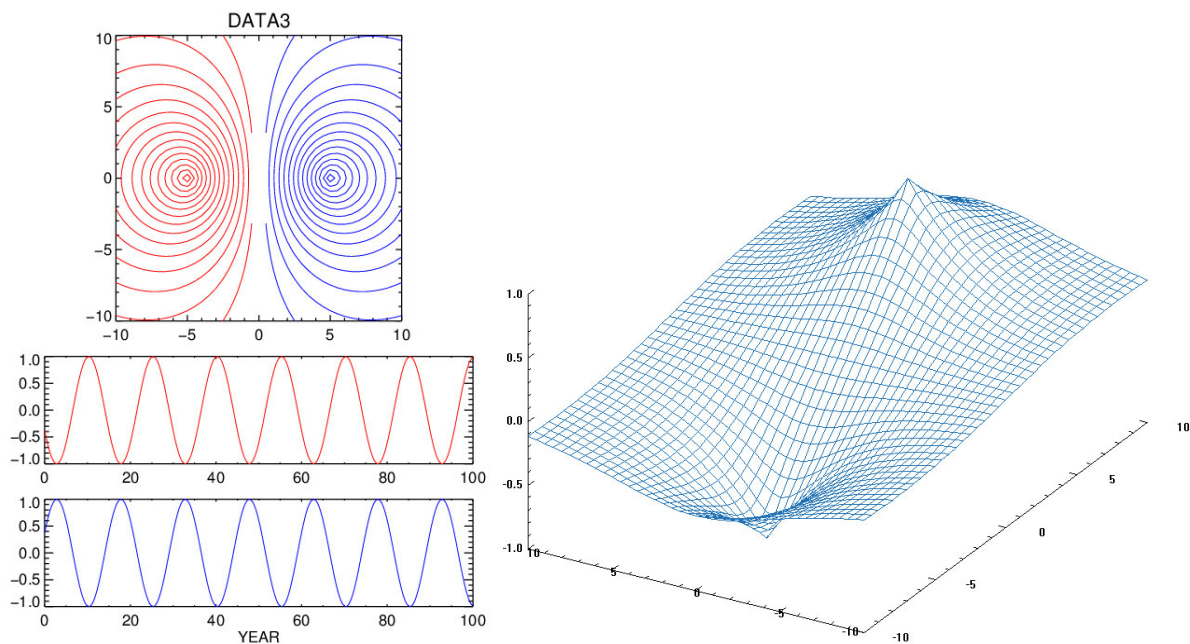


Figura 3.3. Dipolo Este-Oeste

El cuarto producto “DATA4” tiene tres centros de acción en la dirección horizontal y oscila con un periodo de 30 años (360 meses). La figura 3.4 muestra su estructura espacial con su serie de tiempo asociada

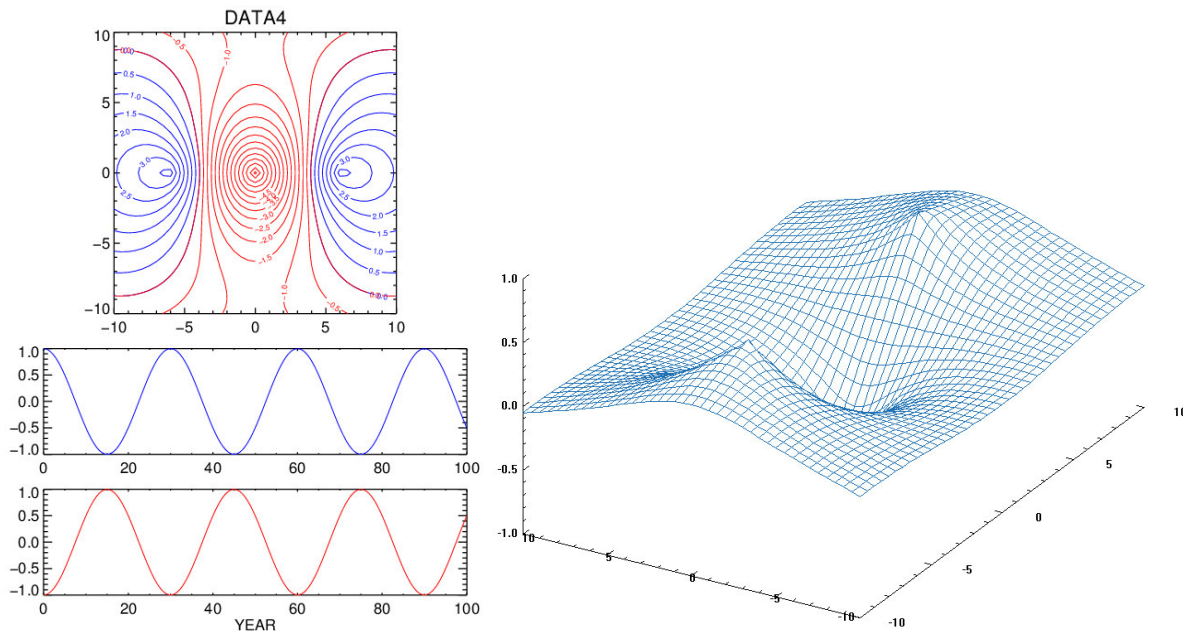


Figura 3.4. Tripolo Este-Oeste

Los primeros 4 productos representan oscilaciones estacionarias. El quinto producto “DATA5” representa un patrón de propagación: Dos centros de acción de signos opuestos que se propagan en círculos alrededor del dominio con un periodo de 10 años (120 meses).

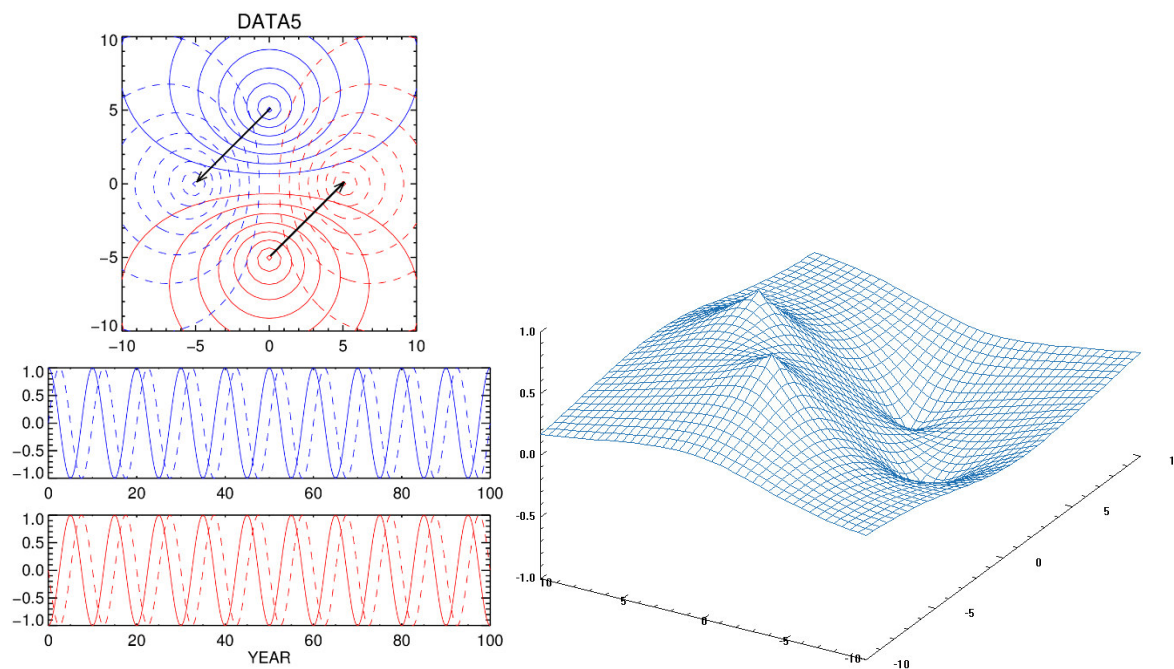


Figura 3.5: Patrón de propagación

Por combinación lineal de dos o más de los conjuntos de los productos sintéticos mencionados, construiremos diferentes campos para analizar cómo funciona el análisis FOE.

Ejemplo 1: campo=Data3 +2*Data2

Los dos primeros patrones del análisis FOE son mostrados en la figura 3.6. El método aisló Data2 como el primer modo (casi 80% de la varianza total) y Data3 como el segundo patrón (20% de la varianza total). Data2 fue claramente el patrón dominante debido a que su amplitud fue multiplicada por 2.

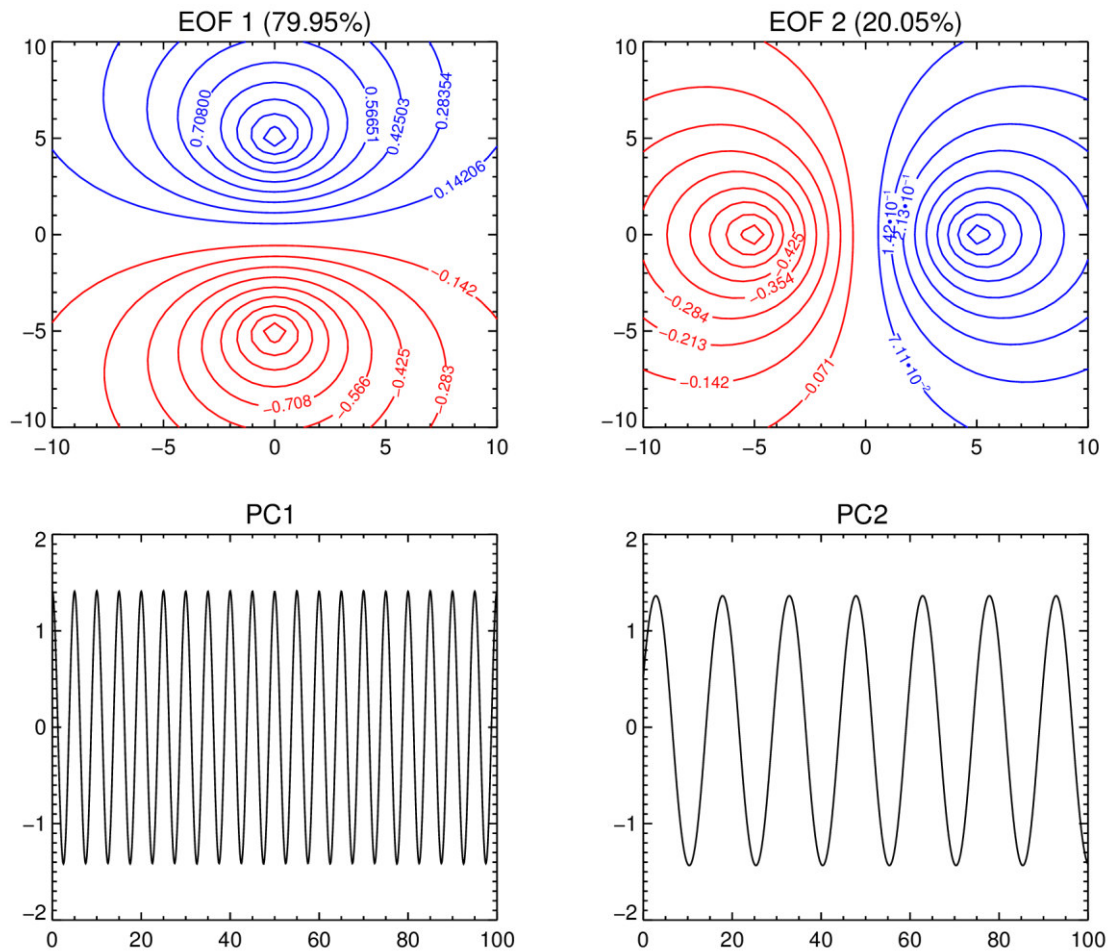


Figura 3.6: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data3 +2*Data2

Ejemplo 2: campo=Data3 + Data2

Los dos primeros patrones son mostrados en la figura 3.7. En este caso ya que ambos patrones son dipolos con amplitudes comparables, el método no es capaz de aislar las dos estructuras esperadas. En lugar de ello el método descompone los datos en dos patrones que explican por igual la varianza total ($\sim 50\%$ cada uno), y cuyos patrones espaciales y temporales son una mezcla de data3 y data2. El método no es capaz de separar las diferentes escalas temporales como nos hubiera gustado en este caso.

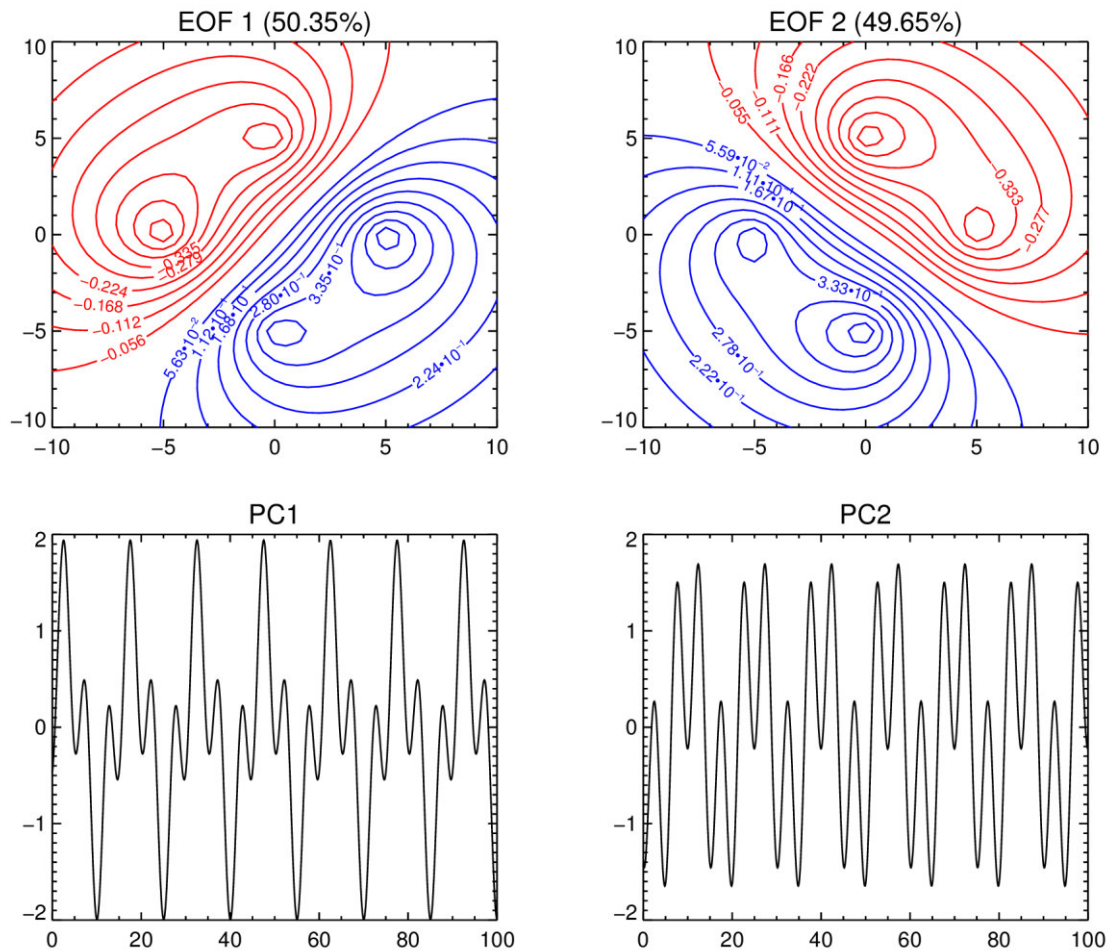


Figura 3.7: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data3 +Data2

Ejemplo 3: campo=data1 + data2

Los dos primeros patrones son mostrados en la figura 3.8. En contraste con lo ocurrido en el ejemplo 2, en este caso el método claramente aísla “Data1” como primer patrón (65.56%) y “Data2” como el segundo patrón (33.44%). La diferencia con el ejemplo 2 es que aunque ambos campos tengan una amplitud similar, uno de ellos es un monopolio y el

otro un dipolo. Esto ilustra el hecho de que el análisis FOE siempre tenderá a aislar como primer patrón la estructura con la escala espacial más grande, en este caso la estructura del monopolo “Data1”.

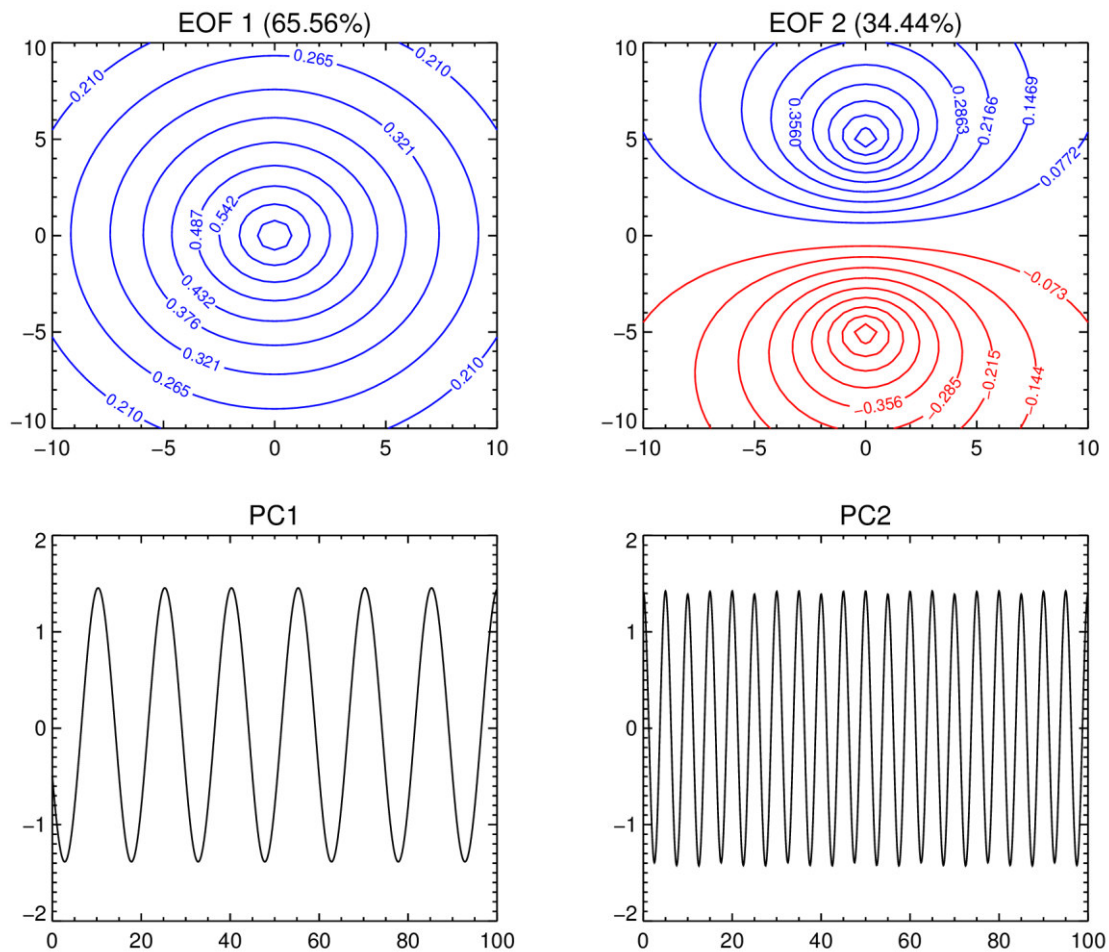


Figura 3.8: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data1 +Data2

Ejemplo 4: campo=data1 + 2*Data2

Los dos primeros patrones son mostrados en la figura 3.9. El hecho de multiplicar “DATA2” por 2 resulta en una descomposición diferente. Ahora DATA2 es el modo dominante y data1 es el segundo. Las varianzas explicadas corresponden al 68 y 32%.

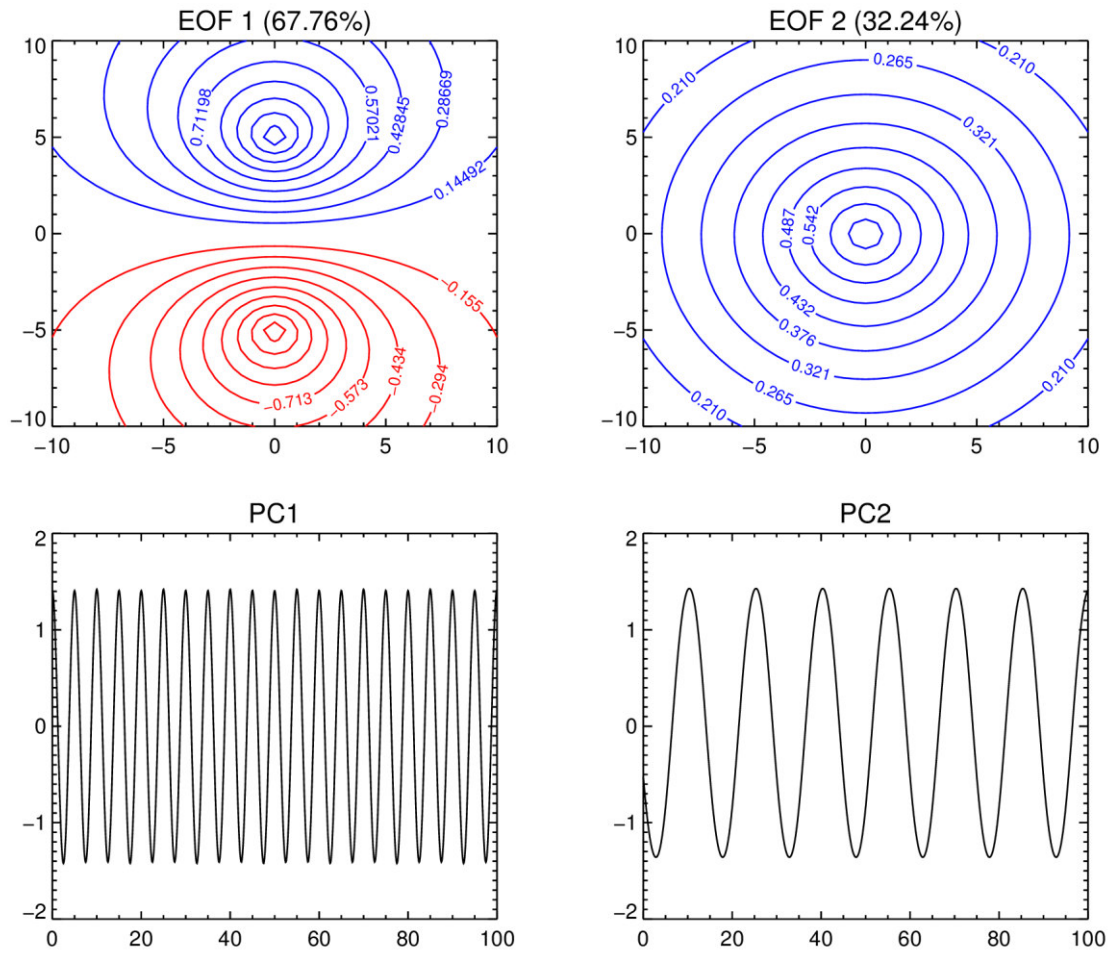


Figura 3.9: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data1 + 2*Data2

Ejemplo 5: campo=Data5

Los dos primeros patrones son mostrados en la figura 3.10. Este ejemplo ilustra el hecho de que la técnica tradicional FOE no es capaz de detectar la propagación de la estructura en DATA5, aunque esta pueda dar una indicación de su existencia. Los dos primeros patrones en esta descomposición están en cuadratura (90 grados fuera de fase) tanto en el espacio como en el tiempo. Este es un intento del método para representar la propagación usando patrones estacionarios. Uno puede imaginar que la señal de oscilación empieza con el patrón del primer modo, luego de un cuarto de ciclo (90°) más tarde se convierte en el patrón del segundo modo, un cuarto de ciclo más tarde se convierte en el patrón del primer modo con signo opuesto, y un cuarto de ciclo después se convierte en el patrón del segundo modo con signo opuesto, lo cual completa el ciclo. Estos patrones poseen una varianza explicada de aproximadamente 50% cada uno.

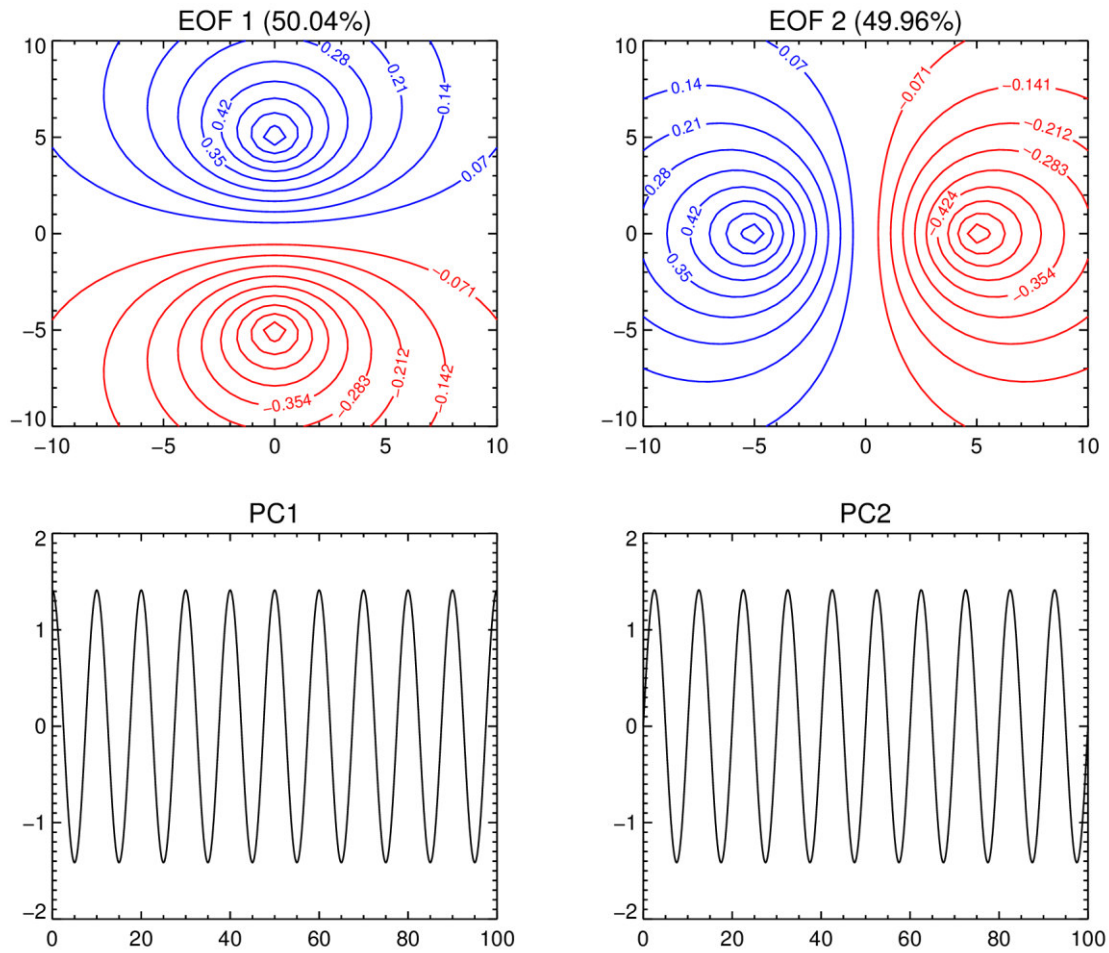


Figura 3.10: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data4

Ejemplo 6: campo= data1 + Data2 + Data4

Los tres primeros modos son mostrados en la figura 3.11. El primer modo claramente aísla el monopolo (DATA1) con el 58.8%, el segundo modo aísla el dipolo (DATA2) con el 29.32%, y el tercer modo aísla los tres centros presentados en la DATA4 con 12.2% de la varianza explicada. Este ejemplo nuevamente ilustra el hecho de que el análisis FOE escogerá como primeros modos aquellos patrones con el mayor número de puntos de grilla con anomalías del mismo signo, es decir preferirá al monopolo por sobre el dipolo y el dipolo por sobre el tripolo, aunque los tres patrones tengan amplitudes comparables.

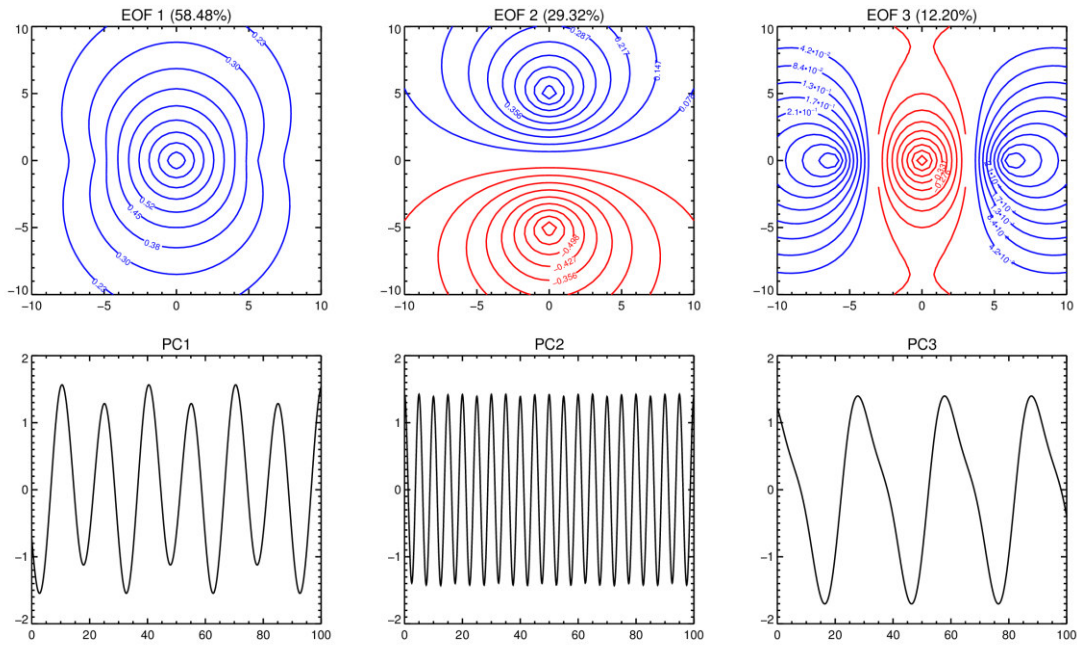


Figura 3.11: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data1+Data2+ Data4

Ejemplo 7: campo= data2 + Data3 + Data5

Los dos primeros modos son mostrados en la figura 3.12. El primer y segundo modo explican una fracción comparable de la varianza (52.41% y 45.59%) y sus patrones

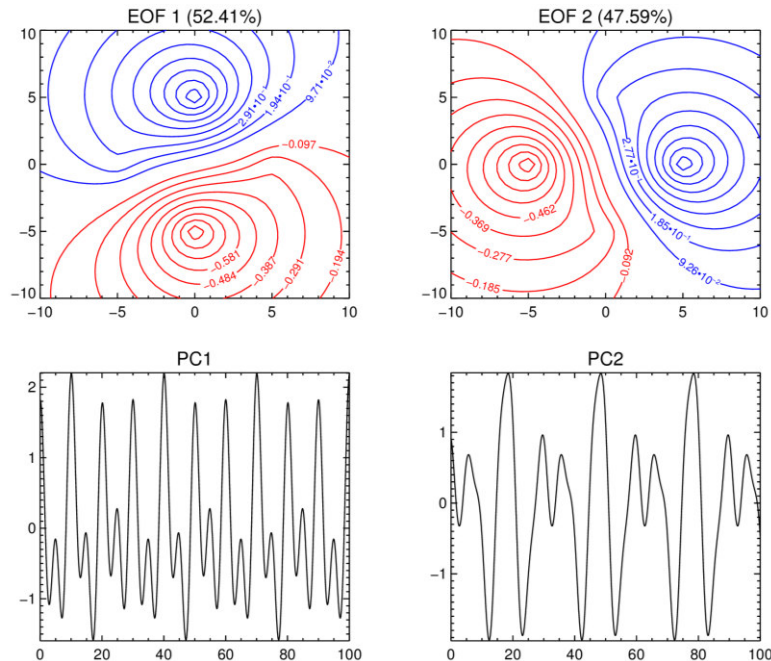


Figura 3.12: Patrones espaciales y temporales del análisis FOE aplicado al campo= Data2+Data3+ Data5

espaciales parecen aislar las estructuras dipolares de DATA2 y DATA3. Sin embargo la evolución temporal del modo 1 muestra una mezcla de las frecuencias de la DATA2 y

DATA5 (5 y 10 años), y el modo 2 muestra una mistura de frecuencias de DATA3 y DATA5 (15 y 10 años). Es decir, la estructura con el patrón de propagación DATA5 se mezcla junto con los otros dos dipolos. El método no logra correctamente aislar los tres patrones en este caso debido a que el patrón espacial de DATA5 se asemeja a los dipolos en ciertas etapas de su propagación.

Ejemplo 8: campo= data1 + Data2 + Data3+data4 + Data5

Los cuatro primeros modos son mostrados en la figura 3.13. La combinación de todos los 5 patrones da como resultado un campo bastante complicado. Sin embargo con la descomposición es posible detectar las principales estructuras espaciales presentes en los datos. Tal como en el ejemplo 7 el modo 1 mezcla el patrón dipolar este-oeste (DATA3) con el patrón de propagación (DATA5), lo cual es reflejado en la mixtura de las escalas temporales (10 y 15 años) en la PC1. El modo 2 mezcla el patrón dipolar norte-sur (DATA2) con otra fase del patrón de propagación (DATA5), también reflejado en la mixtura de las escalas temporales (5 y 10 años) en la PC2. El monopolo de DATA1 está parcialmente capturado en el modo 1, y es responsable de la estructura dipolar sesgada del FOE1. La varianza remanente de DATA1 pareciera ser capturado por el modo3, aunque este modo no aísla claramente ningún patrón original. Sin embargo el modo 4 parece capturar claramente el tripolo de DATA4 en el espacio, pero no tanto en el tiempo.

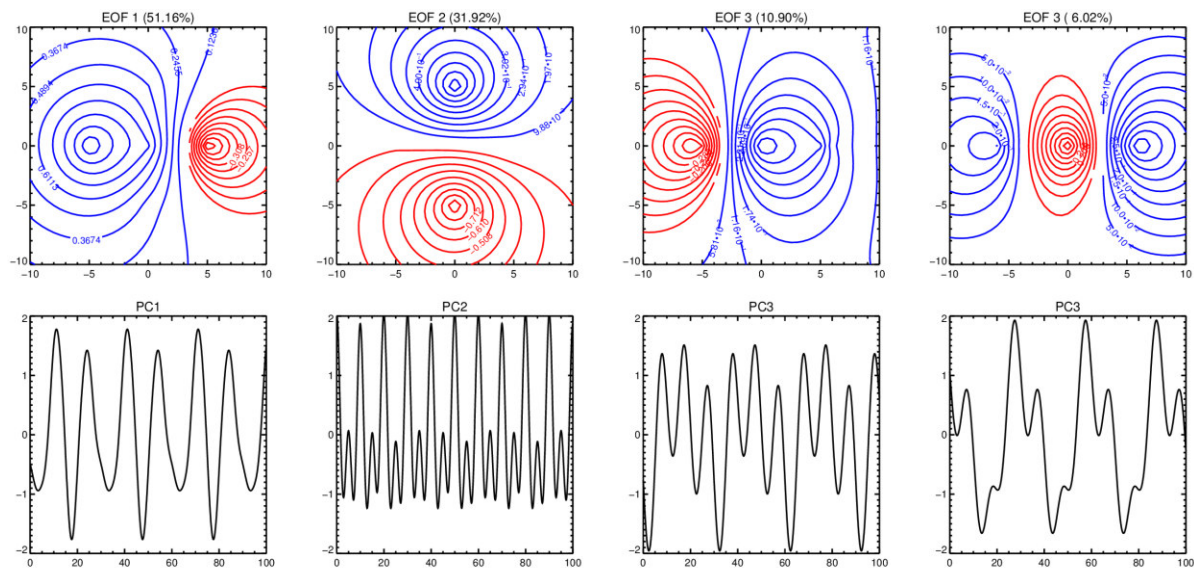


Figura 3.13: Patrones espaciales y temporales del análisis FOE aplicado al campo=Data1+Data2+Data3+Data4+Data5

CAPÍTULO IV

Datos y Metodología

Los productos TSM utilizados en este trabajo son los siguientes:

- **ERSST:** Los datos ERSST (The extended reconstructed sea surface temperature) son datos mensuales de TSM desde 1854 al presente y son derivados del conjunto de datos de ICOADS (International Comprehensive Ocean-Atmosphere Data Set). Estos datos incluyen datos in situ (barcos y boyas), pero no incluyen datos provenientes de satelitales. Los datos faltantes han sido interpolados utilizando métodos estadísticos. La cobertura espacial es 2° latitud x 2° longitud en una grilla global de 89 x 180 píxeles (88° N - 88° S, 0° E - 358°E). Estos datos están disponibles en:
<https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.v5.html>
- **HadISST:** El Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) es una combinación de campos mensuales completos de SST y concentración de hielo marino a nivel mundial para el periodo 1871-presente. HadISST utiliza la interpolación óptima de espacio reducido aplicada a las TSM del Marine Data Bank (principalmente rastreos de barcos) e ICOADS hasta 1981 y una combinación de TSM derivadas de datos satélites e in situ y ajustadas para 1982 en adelante. Estos datos en formato netcdf están disponibles en: <https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>
- **OISST:** Optimum Interpolation (OI) Sea Surface Temperature (SST). Estos datos de 1° de resolución espacial proporcionan campos globales que se basan en una combinación de observaciones de temperatura del océano desde plataformas satelitales e in situ (es decir, barcos y boyas). Los métodos estadísticos (interpolación óptima, OI) se aplican para completar donde faltan valores. Este dato se puede descargar del siguiente link:
<https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>

Nombre	Resolución Temporal	Resolución Espacial	Rango de Años	Formato
ERSST	mensual	2°	1854-presente	Hdf5
HadISST	mensual	1°	1871-presente	netcdf
OISST	mensual	1°	1982-presente	netcdf

Tabla 4.1. Productos globales de TSM

La metodología utilizada para aplicar el análisis FOE a los datos antes mencionados es la siguiente:

- 1) Consideremos un conjunto de N mapas en los tiempos $t = 1 \dots N$, donde cada mapa contiene mediciones del campo ψ en los sitios $p = 1 \dots P$. Por lo tanto, tenemos P series de tiempo $\psi_p(t)$ todas de longitud N , con $P > N$.
- 2) En el análisis de datos geofísicos, con frecuencia se tiene interés en la variabilidad en escalas de tiempo distintas a la escala anual o estacional. En este caso, el primer paso es remover las variaciones estacionales restando el promedio anual climatológico μ_p del campo observado (anomalías). Se puede trabajar con las anomalías o con las anomalías estandarizadas. La estandarización es especialmente relevante cuando se analizan dos o más campos de manera conjunta para asegurar que no domina un campo sobre los otros.

$$X_p(t) = \frac{\psi_p(t) - \mu_p}{\sigma_p} \quad ; \quad \mu_p = \frac{1}{N} \sum_{t=1}^N \psi_p(t)$$

- 3) Se construye la matriz de datos $X_{N \times P}$

$$X = \begin{matrix} & \text{espacio} \longrightarrow \\ \begin{bmatrix} y_1(1) & \cdots & y_P(1) \\ \vdots & \ddots & \vdots \\ y_1(N) & \cdots & y_P(N) \end{bmatrix} & \downarrow \text{tiempo} \end{matrix}$$

- 4) Se construye la matriz covarianza C .

$$C = X * X^T$$

$$C_{N \times N} = \begin{bmatrix} \langle X_1 X_1 \rangle & \langle X_1 X_2 \rangle & \cdots & \langle X_1 X_N \rangle \\ \langle X_2 X_1 \rangle & \langle X_2 X_2 \rangle & \cdots & \langle X_2 X_N \rangle \\ \cdots & \cdots & \cdots & \cdots \\ \langle X_N X_1 \rangle & \langle X_N X_2 \rangle & \cdots & \langle X_N X_N \rangle \end{bmatrix}$$

$$\langle X_i X_j \rangle = \langle X_j X_i \rangle = \frac{1}{N-1} \sum_{t=1}^N X_i(t) X_j(t)$$

La matriz C es simétrica y cuadrada, aún si X no es cuadrada. Si las series de datos en X están normalizadas entonces C será la matriz de correlación en lugar de la matriz covarianza.

- 5) Una vez calculada C se resuelve su problema de eigenvalores-eigenvectores:

$$C * E = E * \Lambda$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_N \end{bmatrix}$$

Generalmente los autovalores en la matriz $\mathbf{\Lambda}$ están acomodados en orden decreciente. Como la matriz de datos \mathbf{X} es real, la matriz de covarianzas \mathbf{C} es definida positiva, lo que significa que todos sus autovalores son mayores o iguales a cero.

Aunque la dimensión de $\mathbf{\Lambda}$ es $N \times N$, típicamente solo los primeros K autovalores son distintos de cero, donde $K \leq \min(N, M)$. Entonces la dimensión efectiva de $\mathbf{\Lambda}$ es de hecho $K \times K$ y por lo tanto solamente se puede determinar K modos o FOEs.

Cada autovalor distinto de cero λ_k está asociado con un autovector columna E^k en la matriz E . Por lo tanto, solamente se usan K autovectores en la descomposición y la dimensión efectiva de la matriz E será $N \times K$. Los autovectores no están correlacionados y cada uno representa el patrón espacial o FOE del modo k .

- 6) La evolución temporal del k -ésimo FOE se obtiene proyectando los datos originales sobre los autovectores, con lo que se obtienen las Componentes Principales:

$$PC^k(t) = \sum_{m=1}^M E_m^k \cdot X_m(t)$$

- 7) Cada autovalor λ_k es proporcional al porcentaje de la varianza del campo observado X que es descrita por el modo k .

$$\% \text{ Varianza} = \frac{\lambda_k}{\sum_{i=1}^K \lambda_i} * 100$$

- 8) El campo original X puede reconstruirse en su totalidad multiplicando cada autovector por su correspondiente PC y sumando los productos:

$$X_m(t) = \sum_{k=1}^K E_m^k \cdot PC^k(t)$$

- 9) Sin embargo, el objetivo de la descomposición FOE es la reconstrucción aproximada, compacta y menos ruidosa, del campo original X usando solamente los H primeros modos, con $H < K$ (H se elige siguiendo los criterios descritos en la sección 3.9):

$$\hat{X}_m(t) = \sum_{k=1}^H E_m^k \cdot PC^k(t)$$

- 10) Si se quiere rotar las FOE's simplemente se utiliza el método VARIMAX (opcional).

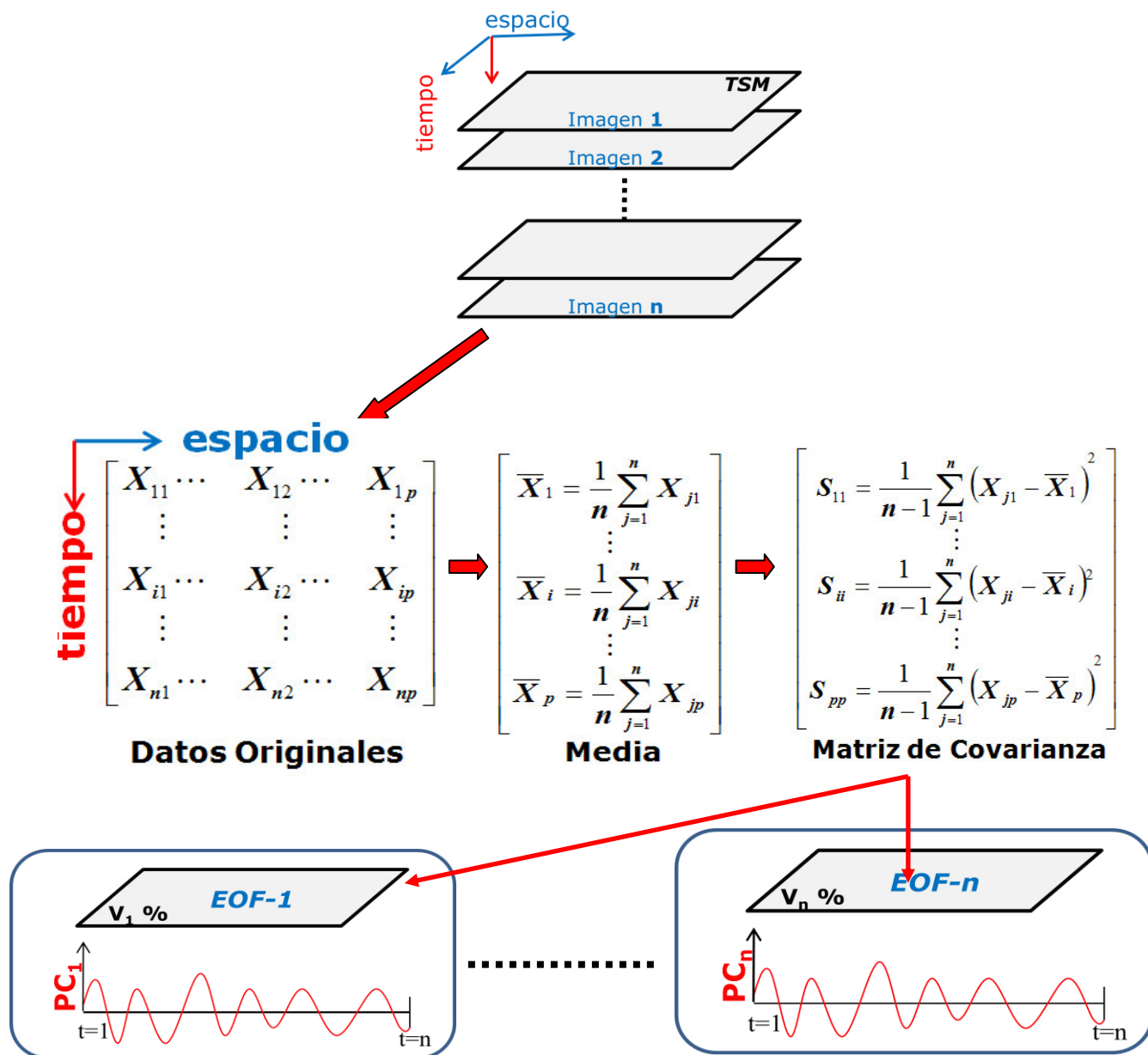


Figura 4.1. Metodología utilizada para obtener las FOE

- 11) Los modos o patrones FOE's tienen tres componentes, los autovalores, los autovectores (patrones espaciales) y las PC's (patrones temporales). Los autovectores son usualmente de interés principal ya que ellos contienen la información de cómo el conjunto de datos se encuentra espacialmente organizado. Existen esencialmente tres formas distintas de presentar los patrones espaciales:
- $\sqrt{\lambda_j} e_j$: Generalmente se presentan los autovectores como patrones espaciales. Sin embargo los autovectores no tienen unidades, por lo que es instructivo presentarlos

de forma escalada (ver sección 3.9). Después de multiplicarlos por la raíz cuadrada de su autovalor asociado estos patrones espaciales ahora ya poseen unidades (por ejemplo °C si se está trabajando con temperaturas) y pueden ser comparados más fácilmente con los datos originales que se utilizaron para el análisis.

- b) ρ_{y,a_j} : En lugar de presentar los autovectores se presentan mapas de correlación entre las componentes principales " a_j " y los datos originales " y ". Estos mapas pueden ser bastante diferentes a las presentaciones con autovectores si el campo de varianza de los datos no es muy homogéneo.
- c) ρ_{y,a_j}^2 : Un patrón que tiene por ejemplo una varianza explicada de 70% no explica 70% en todos los puntos del dominio espacial, sino solo donde la amplitud del autovector es más grande comparada a la desviación estándar. El campo de correlación al cuadrado es el campo de varianza explicada, lo cual es esencialmente poner los autovalores en una perspectiva local.

12) Los autovalores son usualmente presentados en relación a la varianza total como porcentajes, sin embargo es útil presentar los autovalores integrados frente al número de autovalores para encontrar el número de modos necesarios que expliquen cierta cantidad de varianza.

CAPÍTULO V

Resultados y Discusión

5.1 Variabilidad de la TSM en el Pacífico Tropical

El Pacífico tropical es una región conocida por el Fenómeno El Niño (Philander, 1990). Cada 2 a 10 años, un repentino calentamiento de las aguas costeras ocurre frente a Perú. Como el máximo calentamiento ocurre cerca a la navidad, los pescadores locales llaman a este calentamiento 'El Niño'. Durante el tiempo normal, los vientos ecuatoriales del este arrastran las aguas superficiales fuera de la costa, forzando que las aguas sub-superficiales, frías y ricas en nutrientes afloren a la superficie, por lo tanto la productividad biológica y la pesca es muy rica en las aguas peruanas. Durante El Niño, el afloramiento se detiene repentinamente, haciendo que la productividad pesquera entre en crisis. Un fenómeno El Niño considerable puede producir un calentamiento máximo de 5°C o más en las aguas superficiales del Perú. Algunas veces se desarrolla lo opuesto al fenómeno El Niño, es decir que de manera anómala aguas frías aparecen en el Pacífico ecuatorial. A este fenómeno se le ha denominado La Niña. A diferencia de episodios de El Niño, los cuales se habían documentado ya en 1796, los episodios de La Niña se han documentado recién en las últimas décadas, debido a que las anomalías frías de la temperatura superficial del mar se encuentran mucho más lejos de la costa que las anomalías cálidas, no olvidando además que el fenómeno La Niña no afecta a la pesquería peruana.

Los eventos de El Niño tienen influencia mundial e importantes consecuencias socioeconómicas. El conocimiento sobre los cambios en los comportamientos de El Niño es importante para la comprensión y la predicción de El Niño y sus impactos. Los eventos clásicos de El Niño se asocian con una anomalía cálida máxima que ocurre en el Océano Pacífico ecuatorial oriental, conocido como El Niño canónico o el Niño del Pacífico Este (El Niño-PE). En las últimas décadas, ha habido frecuentes apariciones de un nuevo tipo de El Niño con una anomalía cálida máxima en el Pacífico ecuatorial central conocido de diversas maneras tales como El Niño del Pacífico central [Kao y Yu , 2009; Yu y Kim, 2010]; El Niño de la piscina cálida [Kug et al., 2009, o El Niño Modoki [Ashok et al.,

2007]. Estos dos tipos de El Niño tienen diferentes teleconexiones e impactos climáticos [Weng et al., 2009; Kim et al., 2009; Ashok y Yamagata, 2009]. Un estudio realizado por Yeh et al., 2009 ha sugerido que los eventos de El Niño del Pacífico Central (El Niño-PC) podrían ocurrir con mucha más frecuencia en los escenarios proyectados de calentamiento global. Además utilizando los índices históricos de El Niño (el índice Niño3 y el índice Niño4) y los datos ERSST para distinguir dos variaciones de El Niño durante el verano austral (diciembre-enero-febrero, DEF) determinó los años en que ocurrieron el El Niño-PE y el El Niño-PC durante el verano. Desde la década de 1850 (ver Tabla 5.1), el El Niño-PE ocurrió 32 veces y el El Niño-PC ocurrió 8 veces. Cabe recalcar que los eventos La Niña también se presentan como eventos del Pacífico Central y del Pacífico Este (La Niña 1988-89 y La Niña 2007-08)

	El Niño-PE	El Niño-PC
1870s	1876, 1877	
1880s	1888	
1890s	1896, 1899	
1900s	1902, 1904, 1905	
1910s	1911, 1913, 1914	
1920s	1925	
1930s	1930, 1939	
1940s	1940, 1941	
1950s	1951, 1957	
1960s	1963, 1965, 1969	1968
1970s	1972, 1976, 1979	1977
1980s	1982, 1986, 1987	
1990s	1991, 1997	1990, 1992, 1994
2000s	2002, 2003, 2006	2001, 2004, 2009

Tabla 5.1: Años en que ocurrieron los eventos El Niño-PE y El Niño-PC (Yeh et al., 2009)

Utilizando la información contenida en la tabla 5.1, se generó un compuesto de los años en que ocurrieron los eventos El Niño EP y PC (figura 5.1 y 5.2), observándose que los eventos El Niño-PE están caracterizados por anomalías máximas de TSM en el Pacífico Ecuatorial Este; por otro lado el centro de máximas anomalías para los eventos El Niño-PC está ubicado aproximadamente en los 180°W. Una forma de ver con más detalle estas posiciones es generando una gráfica de la estructura zonal para ambos compuestos, utilizando un promedio entre los 2°N y 2°S (figura 5.3). Esta figura indica claramente que

el centro de anomalías máximas para el El Niño-PC está desplazado hacia el Oeste en comparación a El Niño-PE.

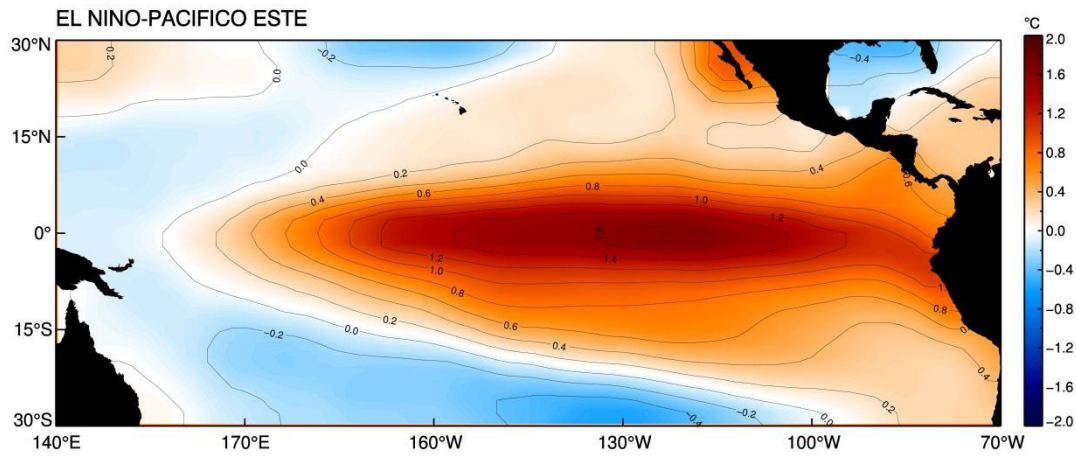


Figura 5.1. Compuesto de las anomalías de TSM (datos ERSST) de los eventos El Niño del Pacífico Este

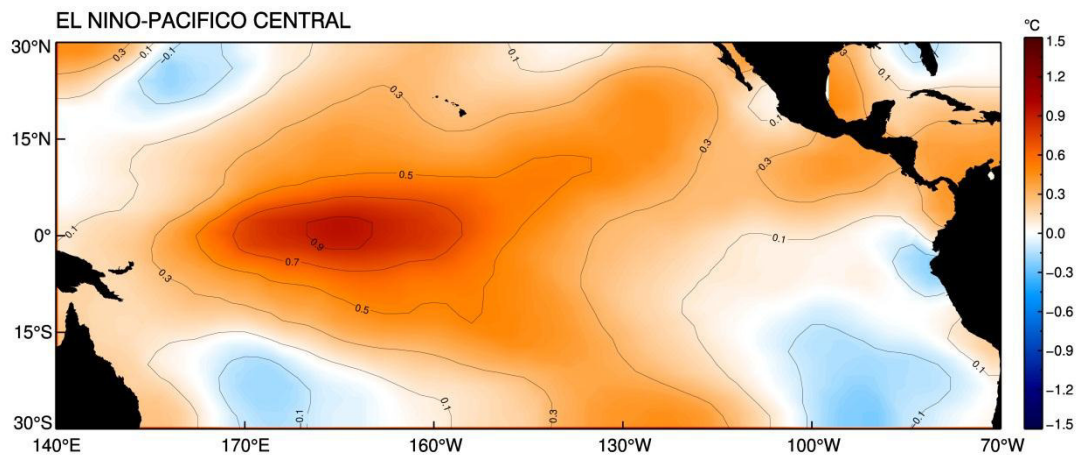


Figura 5.2. Compuesto de las anomalías de TSM (datos ERSST) de los eventos El Niño del Pacífico Central

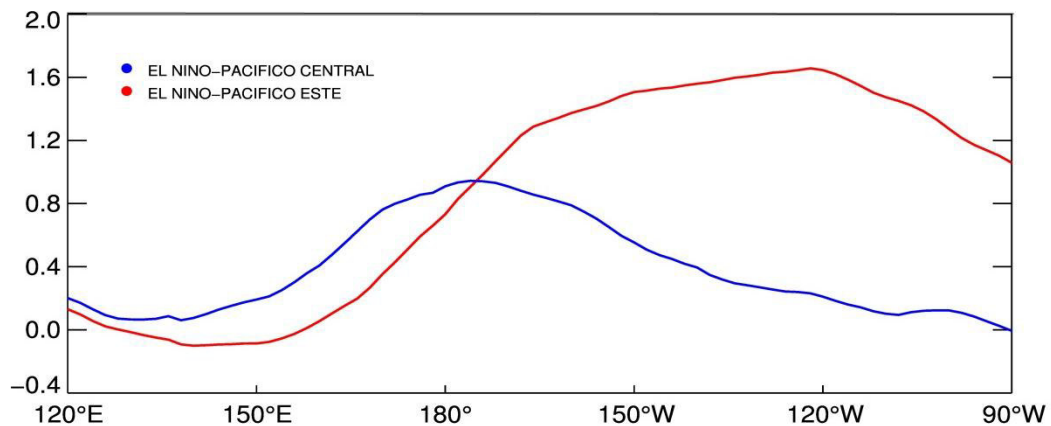


Figura 5.3. Estructura zonal de los compuestos de las anomalías de TSM (datos ERSST) de los eventos El Niño-PE y el Niño-PC promediado sobre los 2°N - 2°S.

Si bien los estudios anteriores han discutido una frecuencia creciente de eventos de El Niño en las últimas décadas, no se han documentado cambios en la magnitud de estos eventos.

Capotondi et al. 2015 señala que los eventos ENSO no son únicamente eventos del Pacífico Central o del Pacífico Este, sino diferentes combinaciones lineales de estos dos tipos de ENSO, los cuales a su vez resultan de diferentes balances de procesos físicos. Para eventos El Niño-PE las anomalías típicamente aparecen en el extremo este del Pacífico durante la primavera y se extienden hacia el oeste durante el verano y el otoño, mientras que en los casos de El Niño-PC las anomalías se extienden desde los subtrópicos orientales hasta el Pacífico Central durante la primavera y verano. Ambos tipos de eventos alcanzan su máxima amplitud durante el verano austral. El inicio y evolución de los eventos individuales de cualquier tipo dependen en cierta medida del momento y la fuerza del forzamiento atmosférico de alta frecuencia que desencadena cada evento.

5.2 Índices de la diversidad de los eventos El Niño

Para entender mejor la diversidad ENSO, varios índices han sido introducidos para identificar diferentes tipos de eventos, con énfasis en la fase cálida del ENSO (El Niño). Los siguientes índices han sido construidos a partir de la TSM. Las definiciones de los tipos de “El Niño” a menudo varían con el método utilizado (ver Capotondi et al. 2015).

- a) **Índice Niño 3-4 (Kug et al. 2009; Yeh et al. 2009)**: Un evento “El Niño” es clasificado como un tipo “piscina caliente” si sus anomalías promedio sobre la región Niño4 ($5^{\circ}\text{S} - 5^{\circ}\text{N}, 160^{\circ}\text{E} - 150^{\circ}\text{W}$) excede una desviación estándar, y excede la anomalía promedio sobre la región Niño3 ($5^{\circ}\text{S} - 5^{\circ}\text{N}, 90^{\circ} - 150^{\circ}\text{W}$). Los eventos “Lengua Fría” son caracterizados por anomalías del Niño-3 que excede una desviación estándar y exceden las anomalías del Niño-4.
- b) **Índice Niño Modoki (EMI) (Ashok et al. 2007)**: Este índice es construido como la diferencia entre las anomalías promedio sobre el Pacífico Central ($10^{\circ}\text{S} - 10^{\circ}\text{N}, 165^{\circ}\text{E} - 140^{\circ}\text{W}$) y las anomalías promedio sobre el Pacífico Occidental ($10^{\circ}\text{S} - 20^{\circ}\text{N}, 160^{\circ}\text{E} - 150^{\circ}\text{W}$) y oriental ($15^{\circ}\text{S} - 5^{\circ}\text{N}, 110^{\circ}\text{E} - 70^{\circ}\text{W}$), para enfatizar la relación fuera de fase entre las anomalías en el Pacífico Central versus las anomalías en el Pacífico Occidental y Oriental.
- c) **Índice EP – CP (Kao y Yu 2009, Yu et al. 2012)**: La regresión de las anomalías de TSM sobre el índice Niño 1+2 (anomalías promedio sobre la región $0^{\circ} - 10^{\circ}\text{S}, 90^{\circ} -$

80°W) es usada para remover la componente de ATSM asociada con el calentamiento del Pacífico Este, y luego un análisis de componentes principales es usado para determinar el patrón espacial y el índice temporal asociado de los eventos **CP**. Similarmente la regresión de las anomalías de TSM sobre el índice Niño-4 es usado para remover la componente de ATSM asociada con el calentamiento del Pacífico Central, y luego un análisis PCA es usado para determinar el patrón e índice de los eventos **EP**.

d) Índice E y C (Takahashi et al. 2011): Este método usa dos ejes ortogonales que son rotados 45° en relación a las componentes principales de la ATSM en el Pacífico Tropical. Las proyecciones asociadas de la TSM sobre estos ejes rotados proporciona el índice E (que representa El Niño del Pacífico Este) y el índice C (que representa el Niño y La Niña del Pacífico Central).

En este capítulo se presenta un análisis que ilustra la aplicación del análisis FOE aplicado a datos de TSM en el Pacífico tropical (30° N - 30° S, 140° E - 70° W). En particular se utilizará el análisis FOE para estimar los índices E y C siguiendo la metodología utilizada por Takahashi et al. 2011, y los índices EP y CP utilizando el método propuesto por Kao y Yu, 2009. Las componentes principales estandarizadas son mostradas en grados centígrados. Los patrones de anomalía de TSM son también mostrados en grados centígrados y fueron calculados mediante la regresión de las anomalías de TSM sobre las series temporales (PC's) normalizadas. Posteriormente para caracterizar de una manera más adecuada los modos de variabilidad se ha correlacionado los patrones temporales con las anomalías de las regiones Niño1+2, Niño3, Niño4 y Niño3+4 (Figura 5.5), los cuales fueron obtenidos del siguiente link: <http://www.cpc.ncep.noaa.gov/data/indices/>

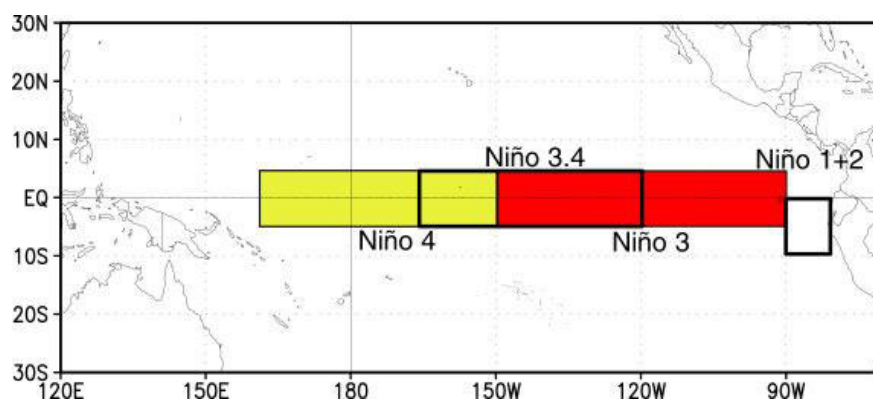


Figura 5.4. Regiones de interés en el Pacífico tropical: **Niño 1+2** (0°-10° S, 80°-90° W), **Niño 3** (5°S -5°N, 170°W -90° W), y **Niño 4** (5°S -5°N, 160°E -150° W). Las anomalías promedio sobre cada uno de estas tres regiones son usadas como índices. Las anomalías de TSM del Niño 3 y el Niño 4 son comúnmente usadas

como índices para los episodios de El Niño/La Niña. (Figura obtenida del Centro de Predicción Climática de la NOAA).

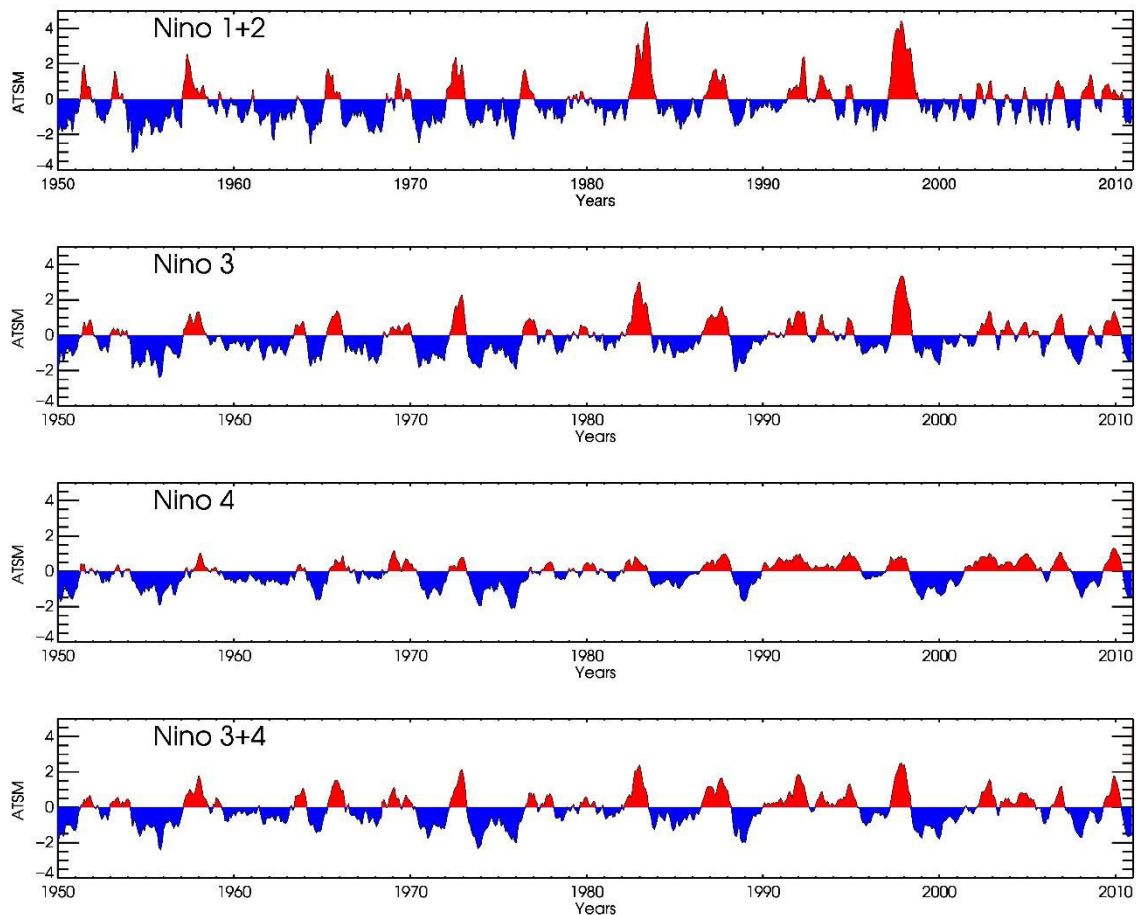


Figura 5.5: Anomalías de las Regiones El Niño

5.3 Patrones de Variabilidad (No Estacionales)

El periodo utilizado para el cálculo de las FOE's será 1950-2010. El periodo base utilizado para calcular las anomalías es 1982-2010. Se usó este periodo base por tres razones: (a) Debido a la alta calidad de las observaciones en la era post-satélite y a la red de boyas TAO, (b) por consistencia en el cálculo de anomalías entre todas las bases de datos de TSM (los datos OISST empiezan en 1982) y (c) por recomendación del World Meteorological Organization (WMO) que sugiere periodos base de aproximadamente 30 años.

El primer patrón o modo de las anomalías de TSM mensuales sobre el Pacífico Ecuatorial claramente representa el patrón ENSO Canónico (Rasmusson y Carpenter 1982), con una extensa región de anomalías de TSM con el mismo signo extendiéndose a lo largo de la

línea Ecuatorial y frente a la costa de Sudamérica y explica un 45.10% de la varianza total de la TSM. Aquí mostramos el patrón dominante basado en anomalías mensuales de los datos HadISST (figura 5.6).

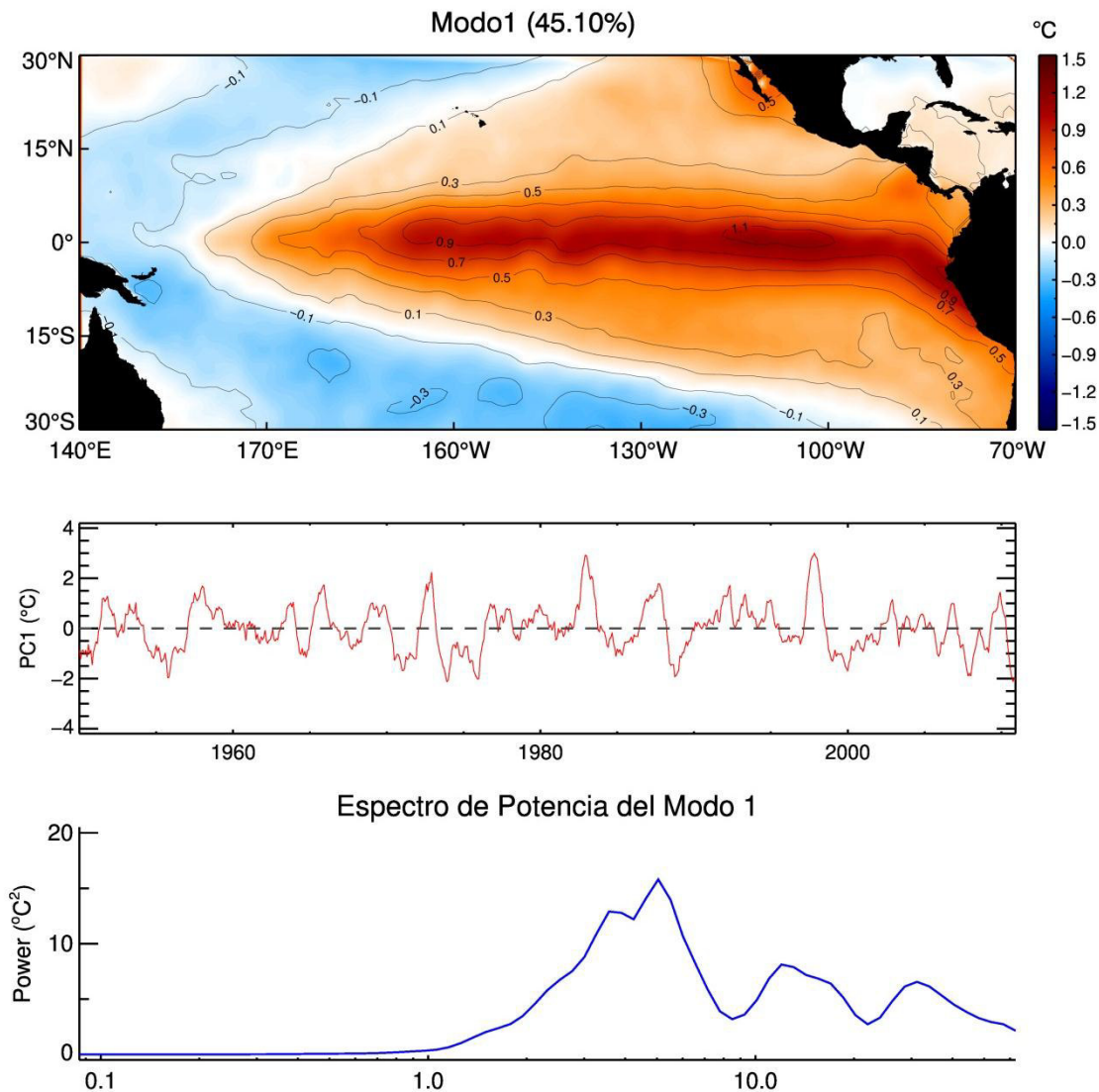


Figura 5.6: Patrón espacial y temporal del modo 1 obtenido de los datos HadISST (1950-2010).

Resultados similares se obtienen utilizando los datos ERSST y OISST. El patrón espacial asociado con la fase cálida del ENSO consiste en anomalías positivas de TSM a lo largo de la zona oriental del Océano Pacífico Ecuatorial, bordeado por anomalías negativas débiles sobre el extremo occidental del Pacífico Tropical y que se extiende en forma de “herradura” hacia el Pacífico Norte y Sur. Anomalías de TSM positivas también son encontradas a lo largo de la costa oeste de Norte y Sudamérica.

Se usa comúnmente el área promedio de anomalías de TSM mensuales en la región 5°N-5°S, 170°-120°W como un índice para representar la variabilidad de la TSM asociada con ENSO, denominado el índice Niño 3-4. El índice de TSM Niño 3-4 es casi idéntico a la serie temporal (PC1) del primer FOE dominante. La serie temporal PC-1 tiene una tendencia lineal positiva y está bien correlacionada con las anomalías de las regiones El Niño (figura 5.7).

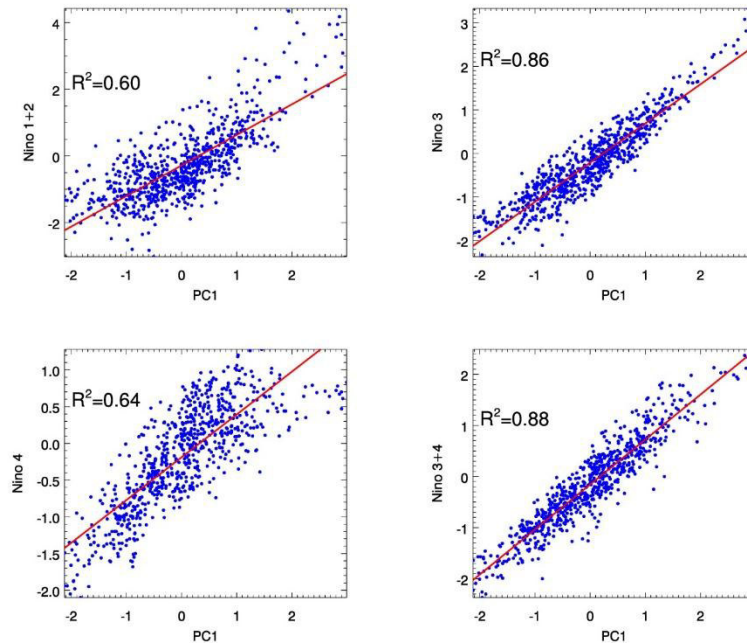


Figura 5.7: Correlación entre el modo 1 (PC1) y las anomalías de TSM de las regiones El Niño.

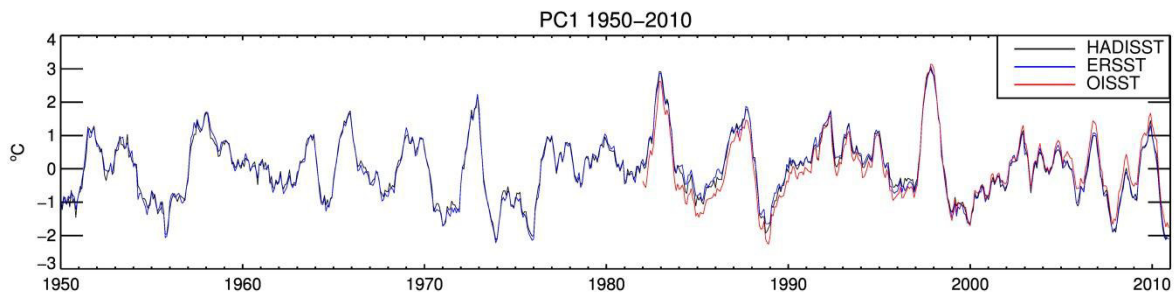


Figura 5.8: PC1 obtenido de los datos ERSST, HadISST y OISST para el periodo 1950-2010.

La series temporales correspondientes a la PC1 para las bases de datos ERSST, HadISST y OISST (figura 5.8) exhiben series irregulares de eventos cálidos y fríos que normalmente duran alrededor 1-1.5 años y se repiten aproximadamente cada 3-8 años. Los eventos cálidos más grandes tales como los ocurridos en 1982-83 y 1997-98 tienen amplitudes de aproximadamente 2°C - 3°C, mientras que los eventos cálidos más moderados van desde

1°C a 2°C. Eventos fríos tienden a ser más débiles y más duraderos que sus contrapartes cálidas. El espectro de potencia de la PC1 muestra que el rango dominante de periodos es 2.5 – 8 años, con cierta sensibilidad al periodo del registro.

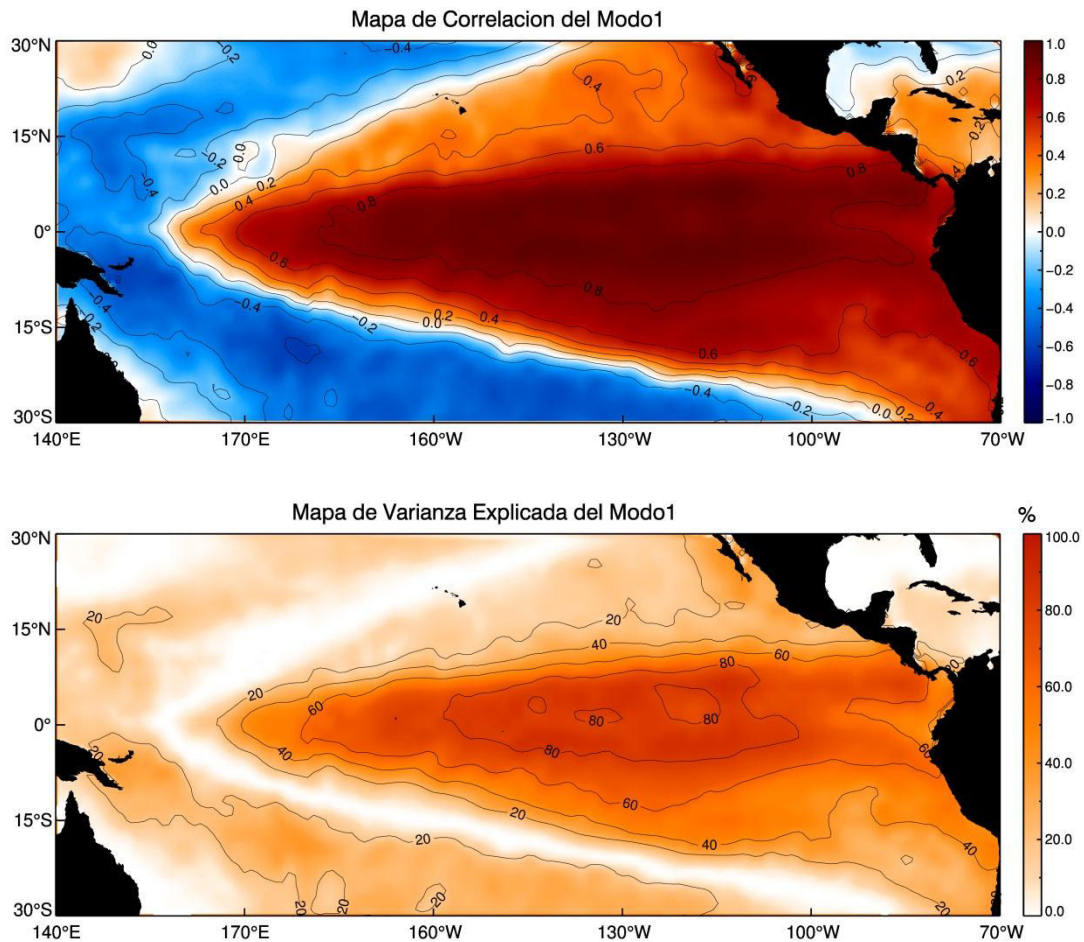


Figura 5.9: Mapas de Correlación y Varianza explicada del primer modo.

El mapa de correlación de la figura 5.9 nos da una perspectiva de las posiciones del dominio en donde existe mayor o menor correlación entre la PC1 y los datos originales de anomalías. El mapa de varianza explicada es simplemente el cuadrado de los valores del mapa de correlación y nos muestra las posiciones que más contribuyen a la varianza conjunta del primer modo que en este caso explica el 45.10% de varianza.

El segundo patrón explica cerca del 10% de la variabilidad de la TSM y captura un ‘**dipolo zonal**’ (figura 5.10). Tanto las anomalías en el este y oeste tienen el mismo signo, mientras que las anomalías en el Pacífico Tropical Central tienen signo opuesto. En altas latitudes

los valores negativos en el Pacífico Ecuatorial Central se extienden hacia el oeste en ambos hemisferios.

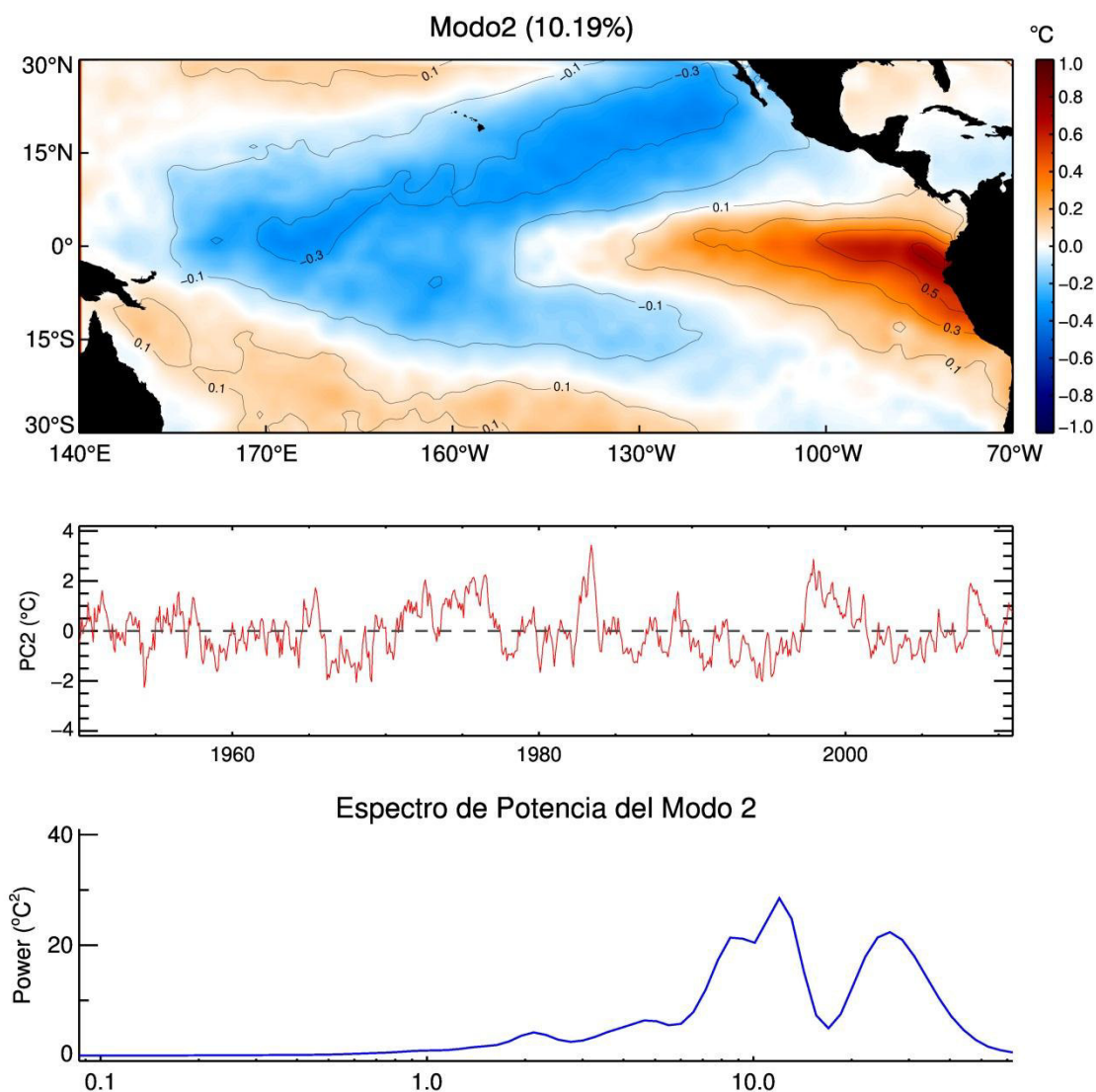


Figura 5.10. Patrón espacial y temporal del modo 1 obtenido de los datos HadISST (1950-2010).

Ashok et al. 2007 que trabajó con datos HadISST en el periodo 1979-2004 se refiere a este modo como El Niño o La Niña “**MODOKI**” (palabra japonesa que significa similar pero diferente) y argumentó un tipo diferente de El Niño. Dicho argumento se refuerza al observar la figura 5.11 en donde las correlaciones de la PC2 con las anomalías de las regiones El Niño son muy bajas. Hay evidencia que este tipo del Niño ha sido más frecuente durante las últimas décadas (Yeh et al. 2009; Lee y McPhaden, 2010). Este segundo modo también recuerda al llamado “Modo del Pacífico Meridional”, que puede

preceder eventos ENSO (Vimont et al. 2003; Chang et al. 2007). La figura 5.10 muestra el patrón temporal PC2 obtenido con las tres bases de datos utilizadas existiendo solo ligeras diferencias en la PC2 del OISST, posiblemente debido a que este solo posee datos desde 1981. Es importante señalar que la escala temporal de este modo es diferente a la del modo1, observándose una fuerte señal decadal que es evidente en el espectro de potencia de la figura 5.10.

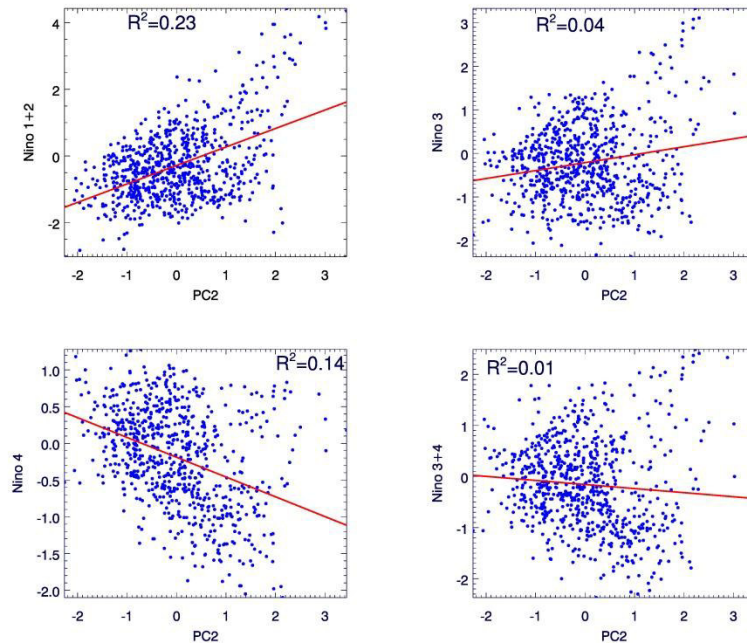


Figura 5.11: Correlación entre el modo 1 (PC2) y las anomalías de TSM de las regiones El Niño.

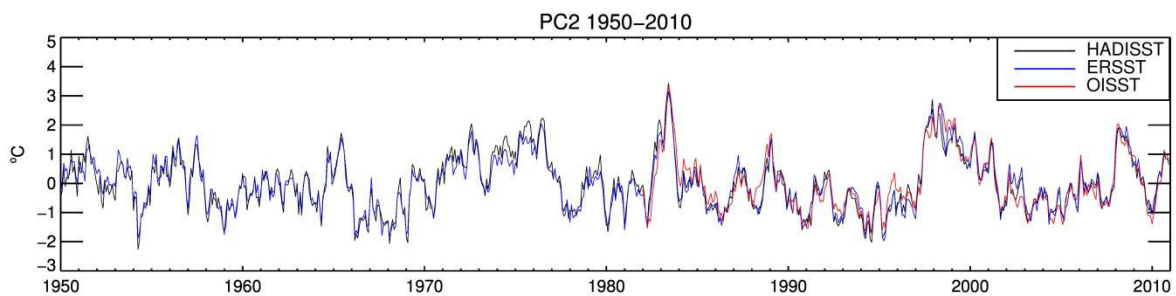


Figura 5.12: PC2 obtenido de los datos ERSST, HadISST y OISST para el periodo 1950-2010.

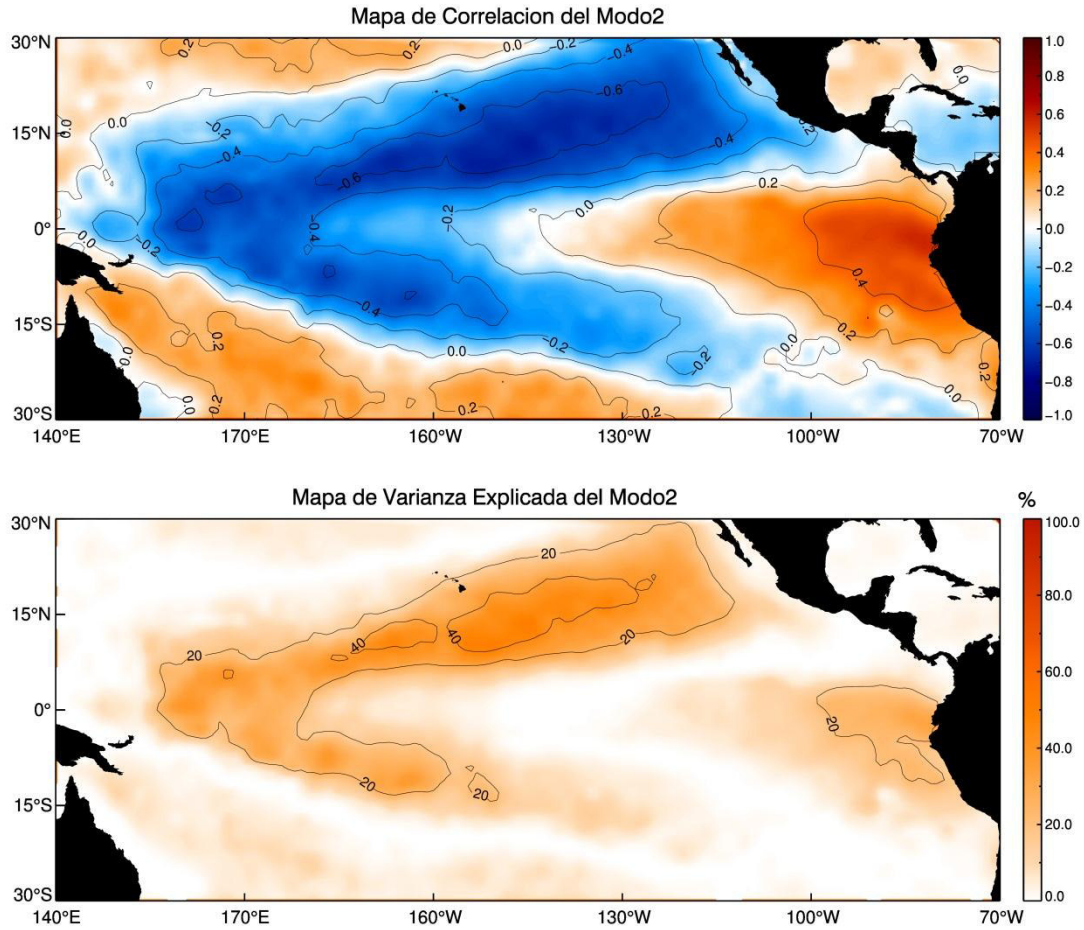


Figura 5.13: Mapas de Correlación y Varianza explicada del segundo modo.

Al igual que en la figura 5.9 se presentan los mapas de correlación y de varianza explicada para el modo 2, siendo el detalle más importante el comportamiento dipolar existente entre el Pacífico Central (incluyendo su extensión al norte y al sur formando una estructura en forma de herradura) y el Pacífico Oriental (principalmente frente a Perú), contribuyendo ambas zonas a la varianza explicada del modo 2 que es de 10.19% (figura 5.13).

En base al patrón espacial y temporal del segundo modo Ashok et al., 2007 derivó un índice denominado El Niño Modoki Index (EMI) que se define como:

$$EMI = [ATSM]_A - 0.5 * [ATSM]_B - 0.5 * [ATSM]_C$$

Los corchetes en la ecuación anterior representan las anomalías de temperatura promedio sobre las áreas indicadas, **A**(165°E-140°W, 10°S-10°N), **B**(110°W-70°W, 15°S-5°N), y **C**(125°E-145°E, 10°S-20°N) respectivamente (figura 5.14).

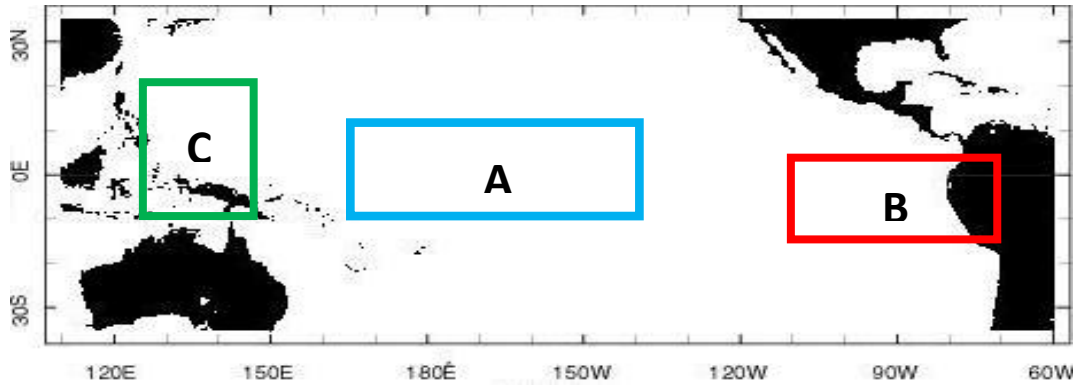


Figura 5.14: Regiones definidas por Ashok et al., 2007 para encontrar el índice EMI

La serie temporal del EMI es mostrada en la figura 5.16 (datos obtenidos de http://www.jamstec.go.jp/frsgc/research/d1/iod/modoki_home.html.en). La correlación entre esta nueva serie temporal y la PC2 es de 0.64 (figura 5.15), lo que comparado con las correlaciones mostradas en la figura 5.9 es significativo. La relación es inversa debido a que hemos considerado el signo opuesto en el cálculo del 2do patrón (es importante recordar que el signo del patrón espacial es arbitrario ya que el producto del 2do patrón y la serie de tiempo PC2 recupera la correcta polaridad), además es más fácil interpretar las relaciones no lineales (ver siguiente sección) si se considera esta elección. Esta correlación muestra que el índice introducido aquí es apropiado para representar los eventos de El Niño Modoki en el Pacífico Tropical.

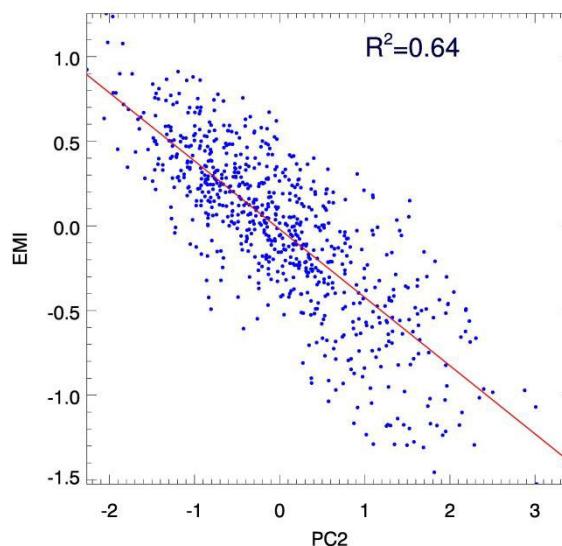


Figura 5.15: Correlación entre el modo2 (PC2) y el índice EMI

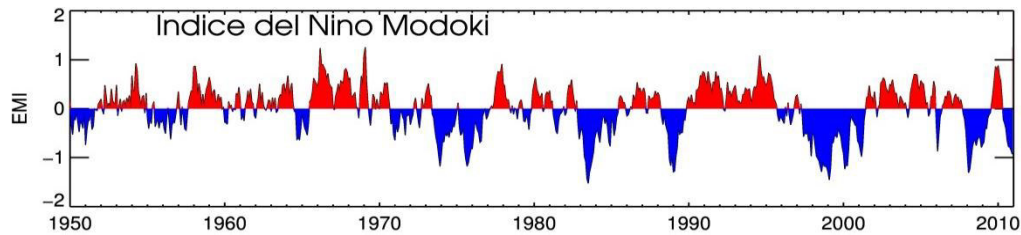


Figura 5.16: Serie temporal del índice EMI

Hasta este momento sólo se han revisado dos modos de variabilidad. El primer modo que explica el 45.10% y el segundo modo que explica el 10.19 %. Ambos suman más del 50% de la varianza total de los datos. ¿Por qué no analizar los siguientes modos? La respuesta se encuentra en el siguiente gráfico que en algunos libros de análisis multivariado lo denominan gráfico de sedimentación (figura 5.17). Este gráfico resulta útil porque permite encontrar de manera práctica el número de modos que son importantes para analizar. Visualmente es fácil notar que estos dos primeros modos están muy alejados del resto. Sin embargo esto es algo subjetivo ya que si cambiamos los datos de temperatura por datos de presión posiblemente la distribución de varianza sea diferente y sea más complicada la tarea de elegir con cuántos modos quedarnos. Por tal motivo recurrimos a uno de los métodos descritos en la sección 2.10 que es la “Regla de North”. En cada punto se ha agregado una barra que corresponde al error de estimación de los autovalores (ver ecuaciones 56 y 57).

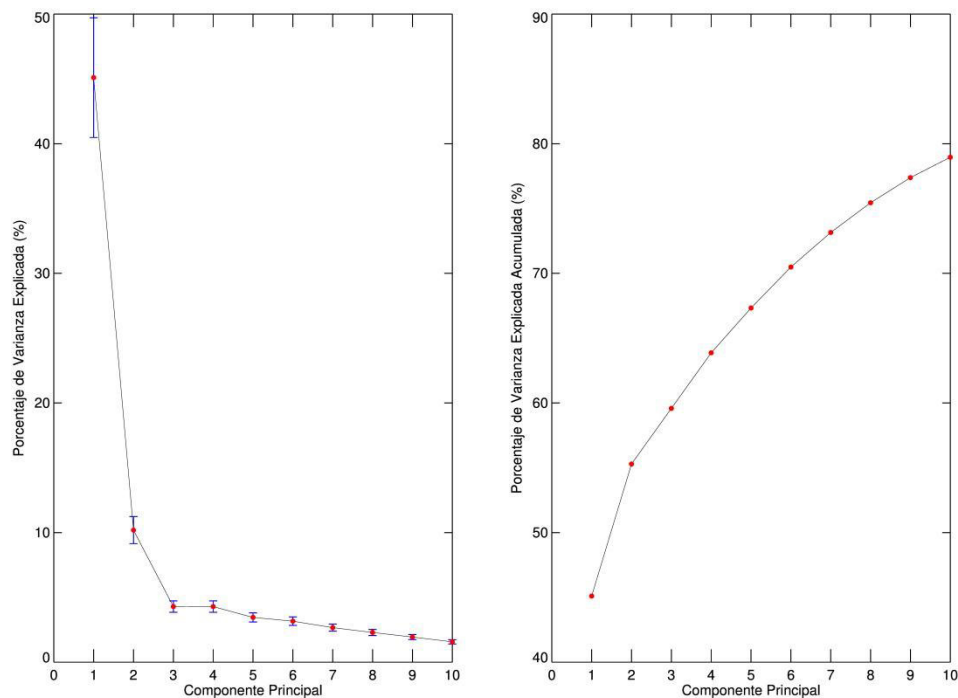


Figura 5.17: Porcentaje de varianza explicada y acumulada por cada modo usando datos HadISST

Según lo explicado en la sección 2.10 si la distancia entre un autovalor y el autovalor más cercano a éste es menor o igual que el error de estimación, entonces estamos ante el caso de autovalores degenerados, lo que en términos prácticos significa que un patrón de variabilidad en un determinado dominio espacial ha sido compartido entre estos dos modos (casi lo mismo que ocurrió al encontrar las FOE's en el ejemplo2 de los datos sintéticos). Si observamos el gráfico 5.17 vemos que el primer modo está bastante separado del segundo, el segundo modo bastante separado del tercero, pero este tercer modo ya no lo está del cuarto, es decir la diferencia entre sus autovalores es comparable a la longitud de sus barras de error, lo cual es una señal de que no debemos tomarlos en cuenta para este análisis.

5.4 Estimación de los patrones correspondientes al Pacífico Central y Este

Como las principales varianzas explicadas por los primeros modos están bien separados es razonable esperar que los modos 1 y 2 representen diferentes modos de variabilidad climática. Sin embargo tal como se explicó en la sección 1.3, uno de los principales problemas del análisis FOE es la restricción que tienen los patrones espaciales a ser ortogonales y los patrones temporales a estar no correlacionados entre ellos. Quizá el primer modo se salve de este tipo de problemas, sin embargo los demás modos estarán sujetos a construcciones artificiales en sus estructuras espaciales. Lo que a continuación se describe es una manera analítica de encontrar el índice que mejor describa los patrones de variabilidad de la TSM.

Con respecto a las diferencias entre ENSO y Niño Modoki se han llevado a cabo importantes investigaciones para identificar, describir y comprender estos tipos de El Niño, lo que ha provocado debates sobre si realmente existen dos modos distintos de variabilidad, o si el ENSO puede describirse más acertadamente como un continuo diverso.

Ashok et al. 2007; Kao y Yu 2009; Kug et al. 2009 documentan las características distintivas asociadas con Modoki, o eventos cálidos del Pacífico Central que no se replican durante eventos ENSO del Pacífico Oriental (Canónico). Por el contrario otros hipotetizan que el ENSO Modoki es un artefacto estadístico que surge como un modo dominante debido a la no linealidad entre El Niño y La Niña (Monahan y Dai 2004). Esta idea es apoyada por el patrón similar al Modoki que se obtiene de la suma de los compuestos de

TSM de El Niño y la Niña, lo cual es debido a grandes anomalías de TSM en el Pacífico Oriental durante el Niño en relación con La Niña (Hoerling et al. 1997).

Lo que se hará a continuación es precisamente realizar estas operaciones (similar a lo que hizo Hoerling et al. 1997). La no linealidad en el patrón ENSO puede en un primer intento ser resumido mediante la división de las anomalías de TSM en 2 categorías, definidas por dos características: El signo y la intensidad. Los patrones promedio para cada una de las 2 categorías se muestran en las figuras 5.18 - 5.19. Se utilizarán los siguientes criterios para encontrar cada uno de los 2 patrones a partir de datos de anomalía de TSM HadISST (sin tendencia lineal) para el periodo 1950-2010:

- a) Eventos Fuertes El Niño: Para todos los datos con anomalías de TSM del Niño34 $> 1^{\circ}\text{C}$
- b) Eventos Fuertes La Niña: Para anomalías de TSM del Niño34 $< -1^{\circ}\text{C}$

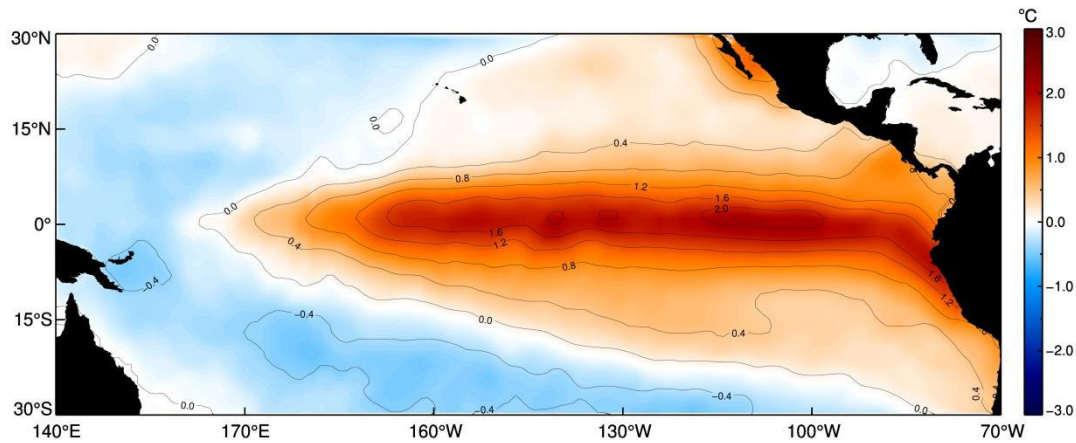


Figura 5.18: Compuesto de los eventos fuertes El Niño obtenido con el criterio descrito en a)

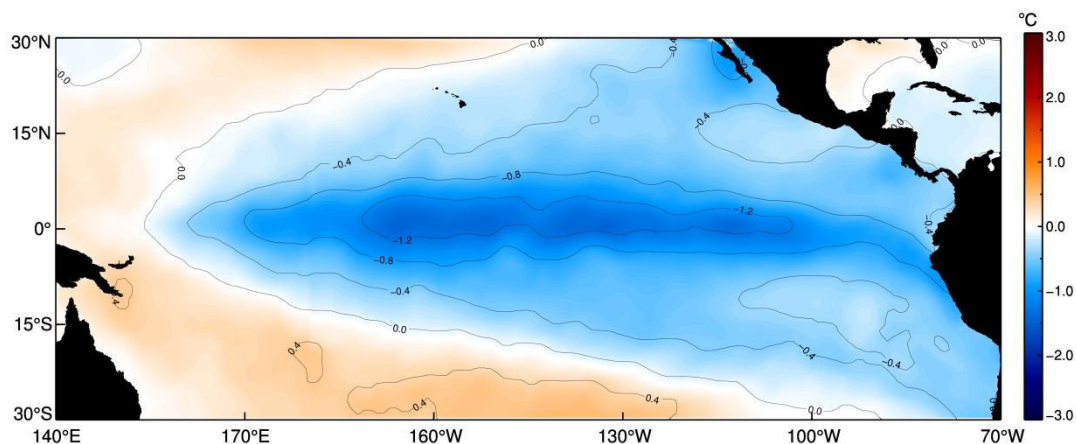


Figura 5.19: Compuesto de los eventos fuertes La Niña obtenido con el criterio descrito en b)

Tenga en cuenta que las anomalías más fuertes de la TSM tienden a ubicarse en el Pacífico central durante los eventos fuertes de La Niña y se centran en el Pacífico oriental durante los eventos fuertes de El Niño. Esta asimetría en los campos compuestos fue notada previamente por Hoerling et al. (1997).

Lo que se muestra a continuación es la suma y resta de los compuestos obtenidos.

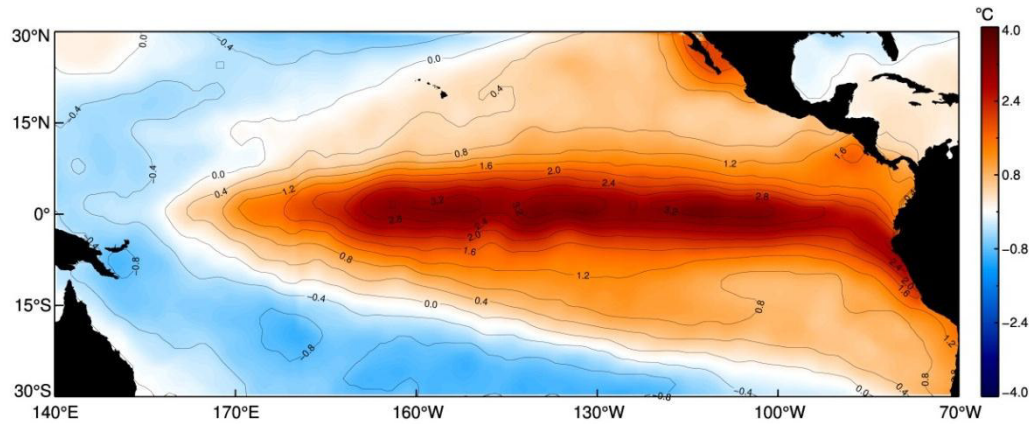


Figura 5.20: Operación eventos fuertes El Niño – eventos fuertes La Niña

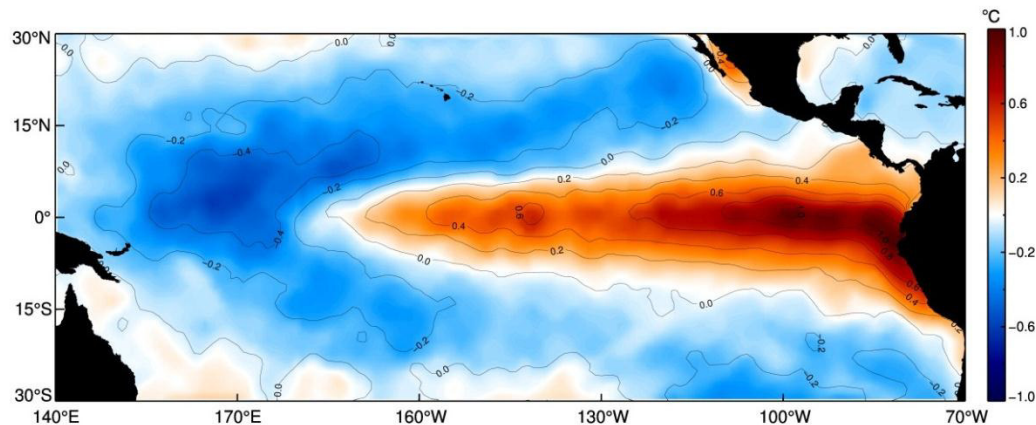


Figura 5.21: Operación eventos fuertes El Niño + eventos fuertes La Niña

Si comparamos las figuras 5.20 y 5.21 con las figuras 5.6 y 5.10 observamos que la diferencia de eventos fuertes El Niño y La Niña aproximadamente reproducen el primer patrón espacial FOE1, y su suma genera el mismo dipolo observado en el patrón espacial FOE2. Es decir que se podría plantear las siguientes relaciones:

$$FOE1 \approx \text{Eventos El Niño}_{Fuerte} - \text{Eventos La Niña}_{Fuerte}$$

$$FOE2 \approx \text{Eventos El Niño}_{Fuerte} + \text{Eventos La Niña}_{Fuerte}$$

Definiendo el Patrón del Pacífico Este a la resultante de la superposición de eventos fuertes El Niño (patrón cuyas mayores amplitudes se ubicarán en la parte oriental del Pacífico) y el Patrón del Pacífico Central como la superposición de eventos fuertes La Niña (patrón cuyas

mayores amplitudes se ubicarán en la parte central del Pacífico), podemos reformular las relaciones planteadas anteriormente y definir nuevas relaciones:

$$\text{Patrón Este} \approx \text{FOE1} + \text{FOE2}$$

$$\text{Patrón Centro} \approx -\text{FOE1} + \text{FOE2}$$

De manera análoga se puede utilizar los patrones temporales (PC's) para plantear lo siguiente:

$$\text{PC_Este} \approx \text{PC1} + \text{PC2}$$

$$\text{PC_Centro} \approx -\text{PC1} + \text{PC2}$$

En donde las componentes principales del Pacífico Este (*PC_Este*) y del Pacífico Central (*PC_Centro*) se pueden estimar a partir de operaciones realizadas sobre las 2 primeras componentes principales.

5.4.1 Estimación de los índices E y C (Takahashi et al. 2011 y Domménget et al. 2013)

Estos autores plantearon que los primeros FOE's de la anomalía de TSM del Pacífico Tropical no describen diferentes fenómenos (es decir el ENSO y el Niño Modoki) sino más bien la evolución no lineal del ENSO, siendo los patrones E y C patrones que representan eventos cálidos extremos en el este y eventos cálidos moderados/fríos en el Pacífico Ecuatorial Central, correspondientes a regímenes con diferente evolución.

A continuación se realizará un análisis de las series temporales PC1 y PC2 analizando primero la gráfica de dispersión entre PC1 vs PC2 (figura 5.22) y posteriormente la evolución temporal de los eventos extremos presentes en estas series.

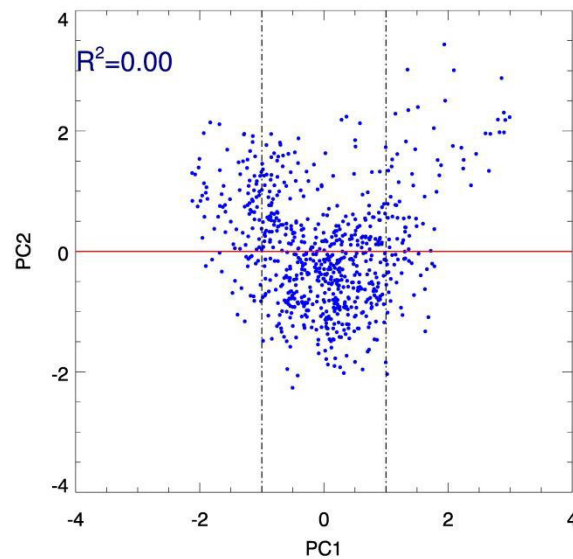


Figura 5.22: Gráfico de dispersión de las series temporales PC1 y PC2

Las series de tiempo correspondientes a estos dos modos temporales no están correlacionadas, pero claramente no son independientes (no forman una nube isotrópica, como se espera para las interacciones lineales). Dicho de otro modo, si estos modos representan modos independientes de variabilidad la dispersión de los puntos debería formar un agrupamiento circular en el origen de coordenadas, sin embargo se observa que presentan una distribución asimétrica, por lo que desde este punto de vista no hay ninguna evidencia de que estos modos representan un nuevo tipo de fenómenos separados del ENSO tal como sostiene Ashok et al. 2007.

La figura 5.22 indica que los valores fuertemente positivos y negativos de la PC1 están asociados con los valores positivos de la PC2. Físicamente, esto describe el hecho de que las anomalías de TSM positivas más fuertes durante un evento promedio de El Niño se encuentran más cerca del límite oriental del Pacífico que las anomalías más frías durante un evento promedio de La Niña.

Esto también se puede encontrar siguiendo los pasos de Dommenget et al, 2013. Para ello se grafican las dos series temporales PC1 y PC2 (Figura 5.23). La franja gris indica el umbral para separar los eventos fuertes de eventos moderados. Las líneas verticales indican 6 eventos extremos de El Niño y La Niña. Se construye un compuesto de la evolución temporal de cada uno de estos eventos usando estos años (figura 5.24), utilizando Diciembre como mes central (que aparece con el número 0 en el eje horizontal). La figura 5.24 indica que la PC2 tiene una covariabilidad no lineal con la PC1: Cada vez que PC1 se encuentra en una fase extrema (evento fuerte El Niño o La Niña) PC2 es positivo. A su vez PC2 tiende a tener valores negativos cuando PC1 está cerca de los valores neutros ($-1 < PC1 < 1$).

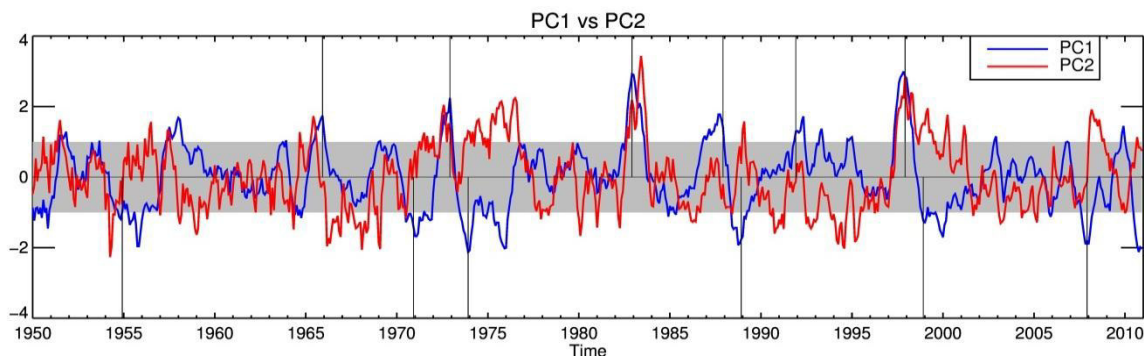


Figura 5.23: Series temporales de las primeras 2 componentes principales (similar a lo presentado por Dommenget et al, 2013).

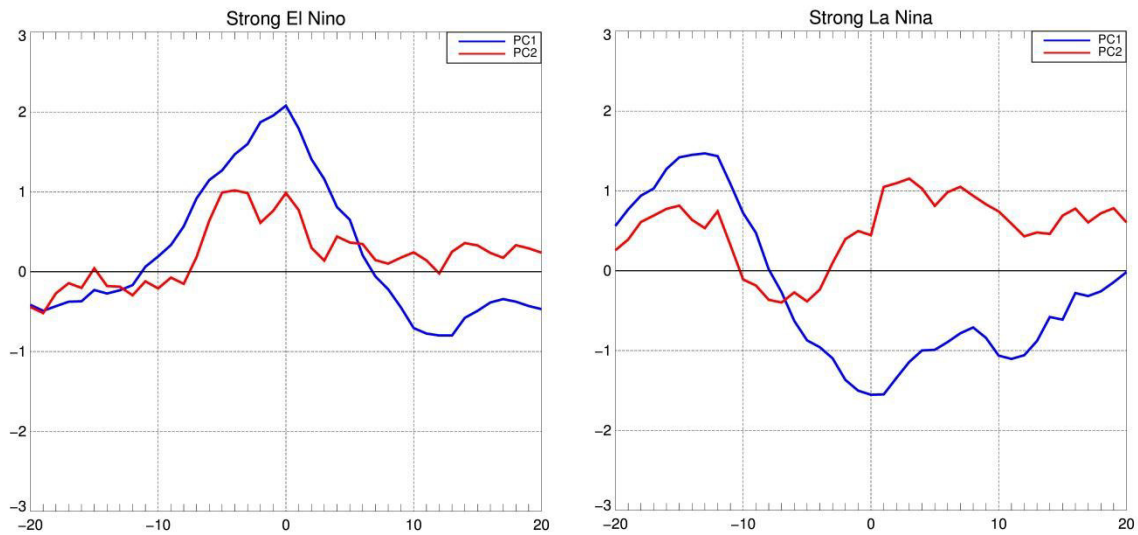


Figura 5.24: Compuestos de la evolución temporal de eventos fuertes El Niño y La Niña (parecido a lo encontrado por Dommenget et al, 2013).

Una vez detectado este comportamiento no lineal, el siguiente paso sera encontrar una relación entre las componentes principales y los índices E y C. Los puntos de dispersión de la figura 5.22 parecen tener dos direcciones ortogonales principales (45° con respecto a los ejes PC1 y PC2) según Takahashi et al. 2011 y Dommenget et al. 2013. Estos ejes son mostrados en la figura 5.25.

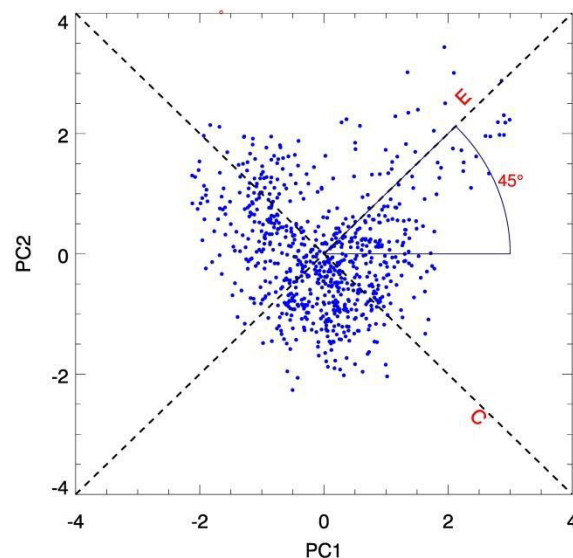


Figura 5.25: Ejes principales de variabilidad de la dispersión de datos PC1 vs PC2

Para rotar la PC1 y PC2 un ángulo de 45° simplemente aplicamos la ecuación A3 del anexo1.

$$E = PC1 \cos 45^\circ + PC2 \sin 45^\circ$$

$$C = -PC1 \sin 45^\circ + PC2 \cos 45^\circ$$

Que es equivalente a:

$$E = (PC1 + PC2) / \sqrt{2}$$

$$C = (-PC1 + PC2) / \sqrt{2}$$

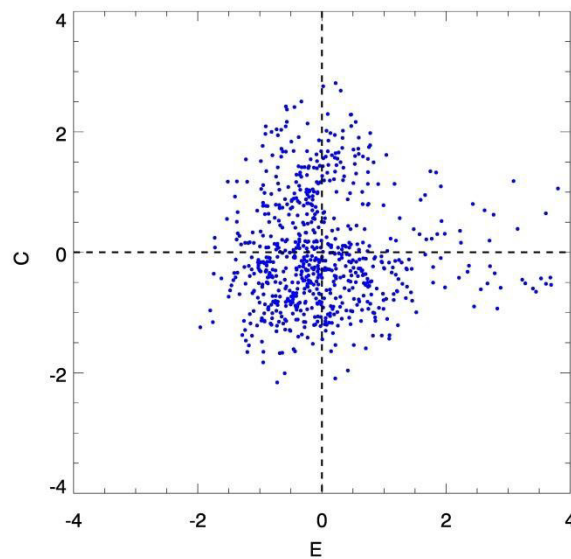


Figura 5.26: Gráfica de dispersión de los nuevos índices obtenidos por rotación (índices E y C)

Una vez obtenida estas nuevas series temporales, se realiza una regresión de los índices E y C con los datos de anomalía para obtener los patrones espaciales E y C.

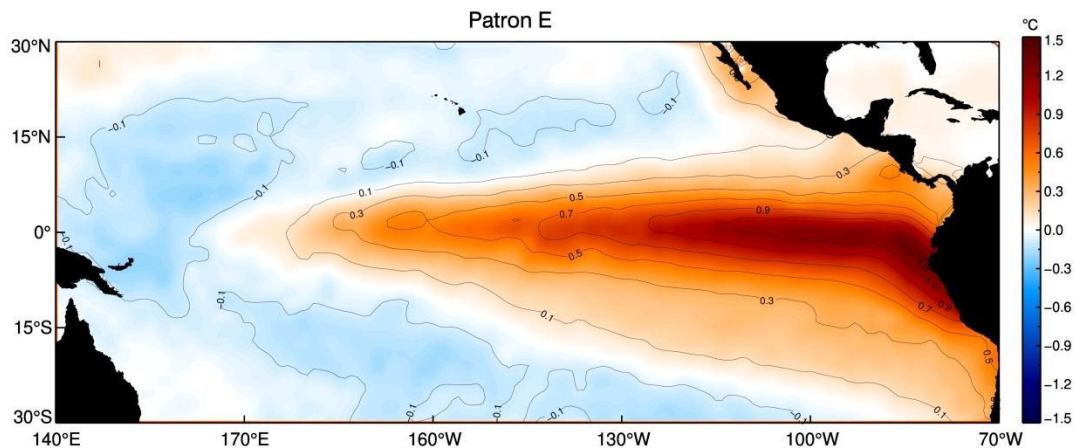


Figura 5.27: Distribución espacial del Patrón E

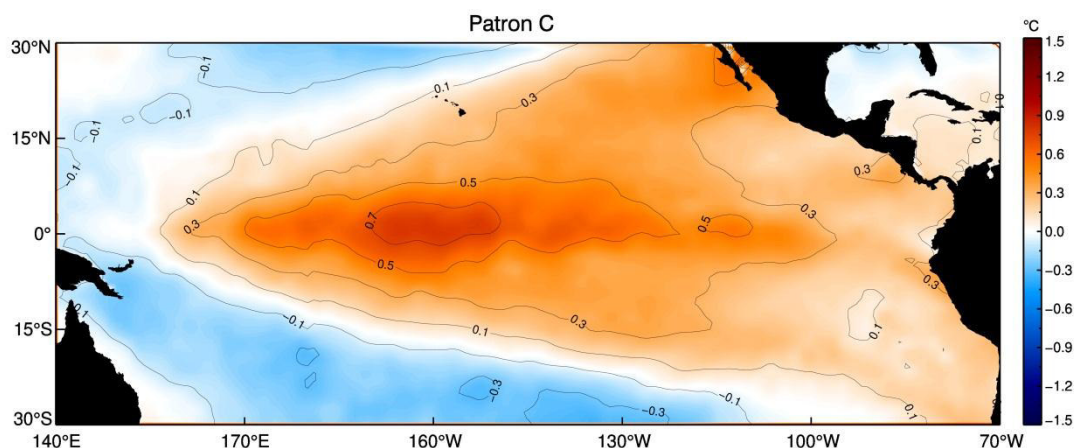


Figura 5.28: Distribución espacial del Patrón C

El patrón E (figura 5.27) es claramente desplazado al este y más confinado al Ecuador que el FOE1, lo cual es consistente con la discusión de la no linealidad de los compuestos de las figuras 5.20 y 5.21. A su vez el patrón C (figura 5.28), el cual es mostrado con el signo inverso, está claramente desplazado al oeste y es más amplio en su extensión meridional que el FOE1. Estos patrones básicamente representan la estructura espacial no lineal de los eventos ENSO de una manera óptima.

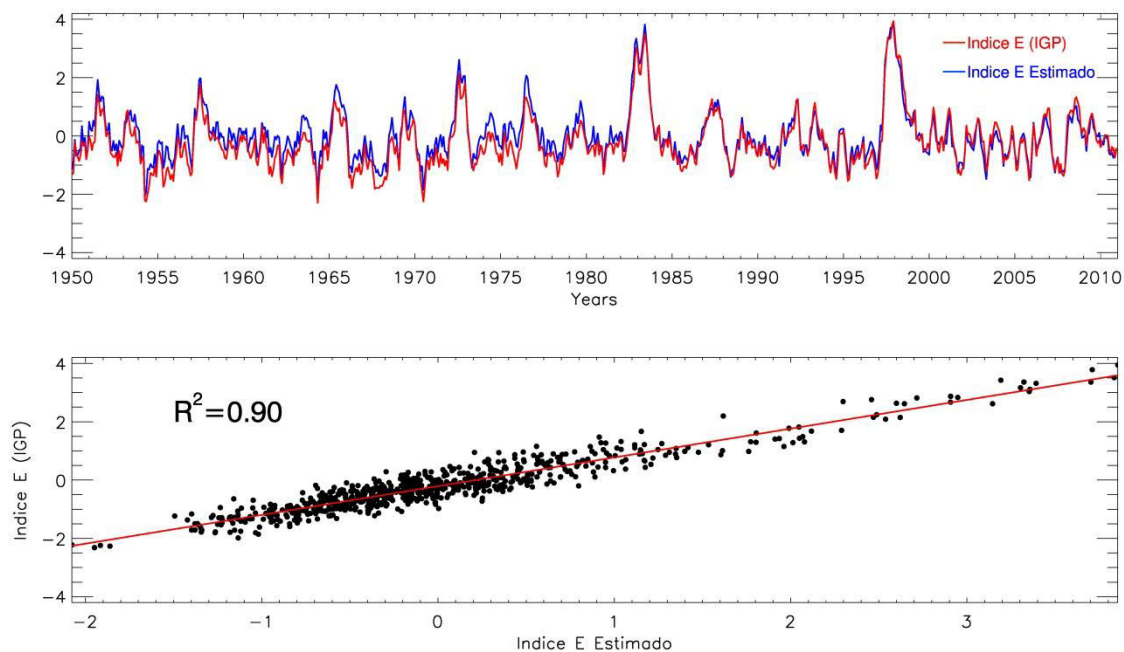


Figura 5.29: Comparación entre el índice E estimado y el índice E calculado por el IGP.

Para evaluar la validez del algoritmo (desarrollado en el lenguaje de programación IDL) con el cual se ha estimado los índices E y C, se realiza una comparación entre estos índices y aquellos producidos en el Instituto Geofísico del Perú (IGP), obteniéndose altas correlaciones (0.9 y 0.86) tanto para los índices E y C (figuras 5.29 y 5.30). Los datos del IGP se obtuvieron del siguiente enlace: <http://www.met.igp.gob.pe/datos/EC.txt>

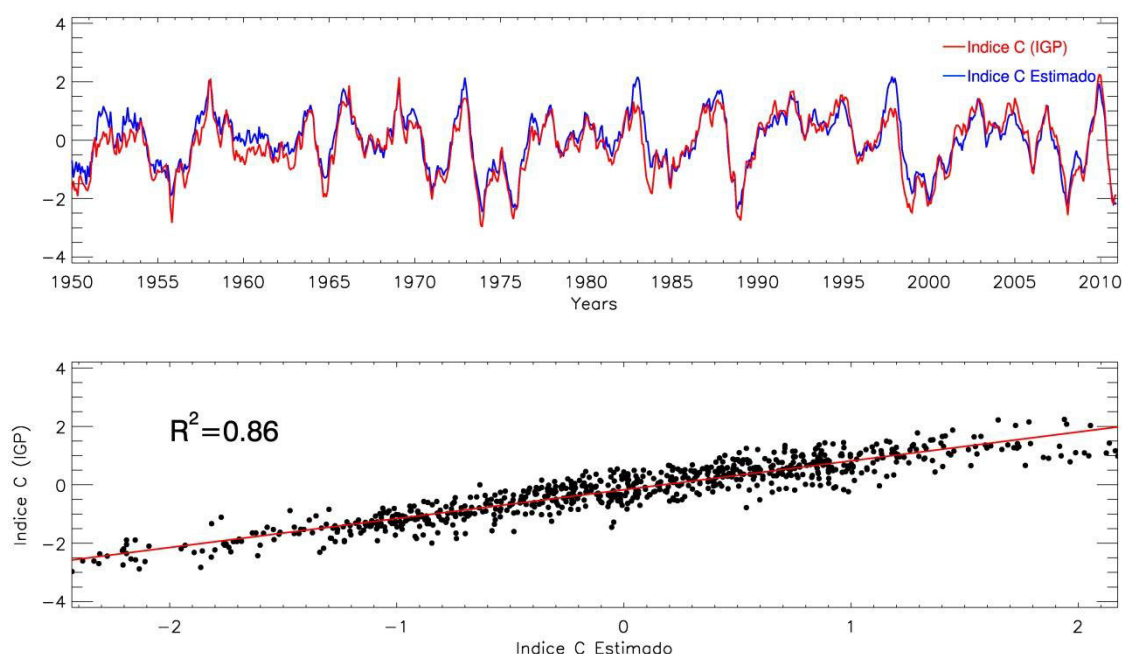


Figura 5.30: Comparación entre el índice C estimado y el índice C calculado por el IGP.

5.4.2 Estimación de los índices EP y CP (Kao y Yu, 2009)

Se utilizan los datos HadISST para contrastar dos tipos de El Niño-Oscilación Sur (ENSO) en el Pacífico Tropical: Un tipo del Pacífico Este (EP) y otro tipo del Pacífico Central (CP). El ENSO-EP tiene sus anomalías ubicadas en el Pacífico Ecuatorial Este (costa de Sudamérica). Este tipo de ENSO está asociado a las variaciones de los vientos superficiales y a la termoclina, y muestra fuerte relación con el Océano Índico Tropical. Por el contrario en el tipo ENSO-CP las anomalías de TSM y vientos superficiales están ubicados en el Pacífico Central. Este tipo de ENSO está menos relacionado con la dinámica (variaciones) en la termoclina y puede estar más influenciado por forzantes atmosféricos. Esta tiene una fuerte teleconexión con el Océano Índico sur (Kao y Yu, 2009). Para encontrar los índices

EP y CP se aplica un método que combina los EOF con el análisis de regresión para separar estos tipos de ENSO (el procedimiento se describe en la sección 5.3c).

El tipo EP-ENSO está caracterizado por anomalías que se extienden desde la costa de Sudamérica en el Pacífico Central a lo largo del Ecuador. Las anomalías están principalmente contenidas en las regiones Niño1+2 y Niño3. Esta estructura es similar a la estructura del EOF principal obtenido de las anomalías de TSM originales (Figura 5.5), pero menos extensa en el Pacífico Central (ver figura 5.31).

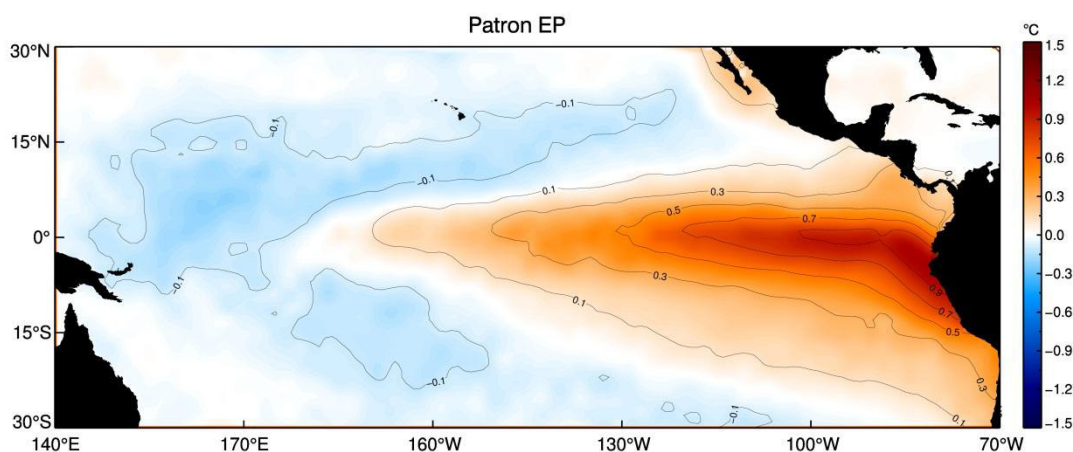


Figura 5.31: Distribución espacial del Patrón EP

La figura 5.32 muestra la estructura del tipo CP-ENSO, el cual tiene anomalías principalmente confinadas en el Pacífico Central entre 160°E y 120°W, cubriendo las regiones Niño3-4 y Niño 4.

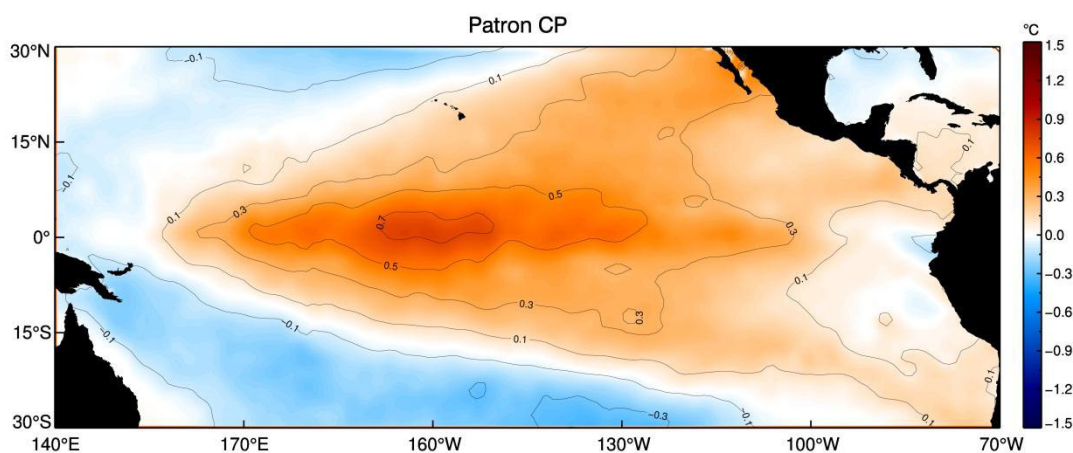


Figura 5.32: Distribución espacial del Patrón CP

Las componentes principales de los patrones EP-ENSO y CP-ENSO son mostrados en la figuras 5.33.

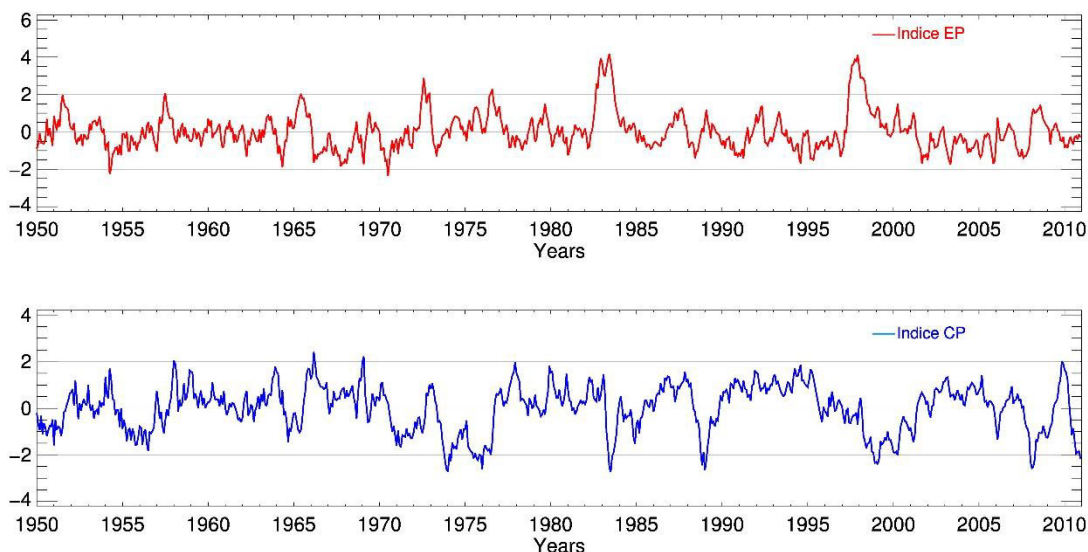


Figura 5.33: Evolución temporal de los índices EP y CP

Se nota que ambas series no son simétricas con respecto al valor cero. Los coeficientes de asimetría de las componentes principales de los tipos EP y CP son 1.3 y -0.47 respectivamente. Los histogramas mostrados en la Figura 5.34 indican que el tipo EP-ENSO está más asociado con eventos fuertes El Niño. Por otra parte el CP-ENSO está más asociado con eventos fuertes La Niña. Esta asimetría espacial fue documentada por varios estudios tales como el de Monahan, 2001 y Hsieh, 2004. No obstante se debe considerar que eventos fríos EP-ENSO y eventos cálidos CP-ENSO también ocurren a menudo.

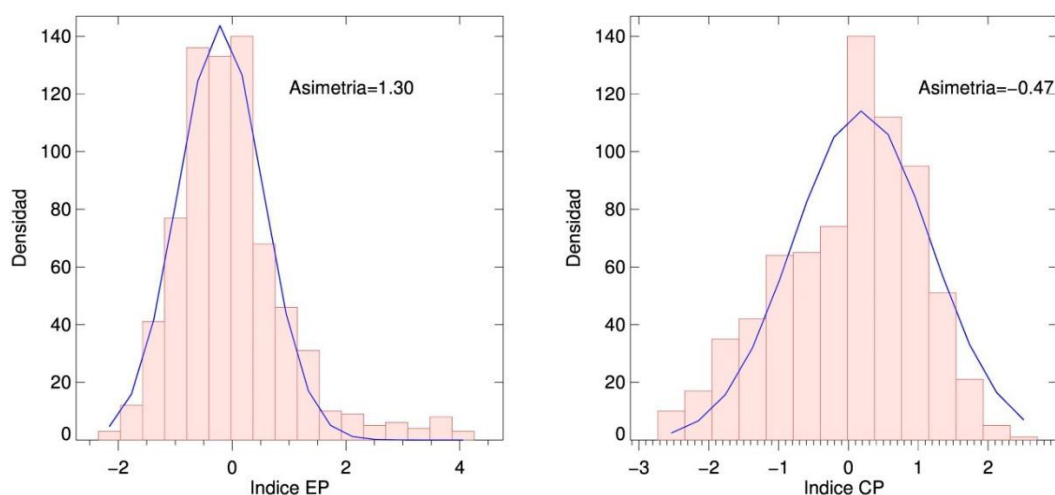


Figura 5.34: Histogramas de los índices EP y CP

Además de las diferencias en su estructura espacial, los tipos EP-ENSO y CP-ENSO también muestran diferencias en sus evoluciones temporales. Las figuras 5.35 y 5.36 muestran los coeficientes de la correlación desfasada (en inglés lagged correlation) entre las componentes principales de los dos tipos de ENSO (EP y CP) y las anomalías de TSM del Pacífico Tropical.

Para el tipo EP-ENSO (figura 5.35), las anomalías empiezan a aparecer en las costas de Sudamérica, se propagan hacia el oeste (llegando hasta el Pacífico Central) y disminuyen su intensidad fuera del Ecuador. Las anomalías más intensas de este tipo de ENSO aparecen cerca a las costas, apareciendo además anomalías débiles y de signo opuesto en el Pacífico Ecuatorial Oeste.

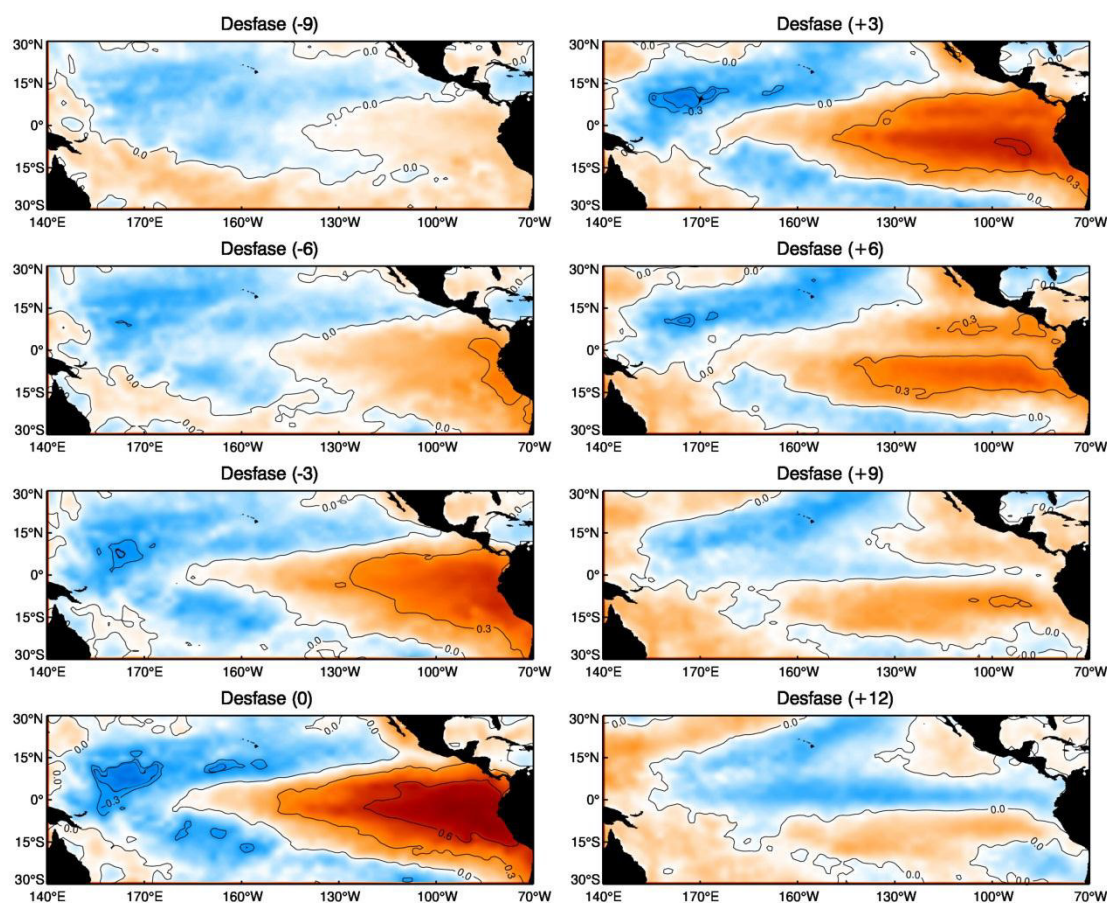


Figura 5.35: Coeficientes de correlación desfasada entre las anomalías del Pacífico Tropical y el índice EP (similar a lo presentado por kao and Yu, 2009).

Para el tipo CP-ENSO (figura 5.36), las anomalías aparecen aproximadamente en los 180°W, desarrollándose hasta adoptar una estructura de anomalía con forma de “herradura”

extendiéndose hacia los subtrópicos en ambos hemisferios, y luego disminuye su intensidad en el Pacífico Ecuatorial Central. El CP-ENSO tiene mayor relación con las anomalías presentes en los subtrópicos que el EP-ENSO.

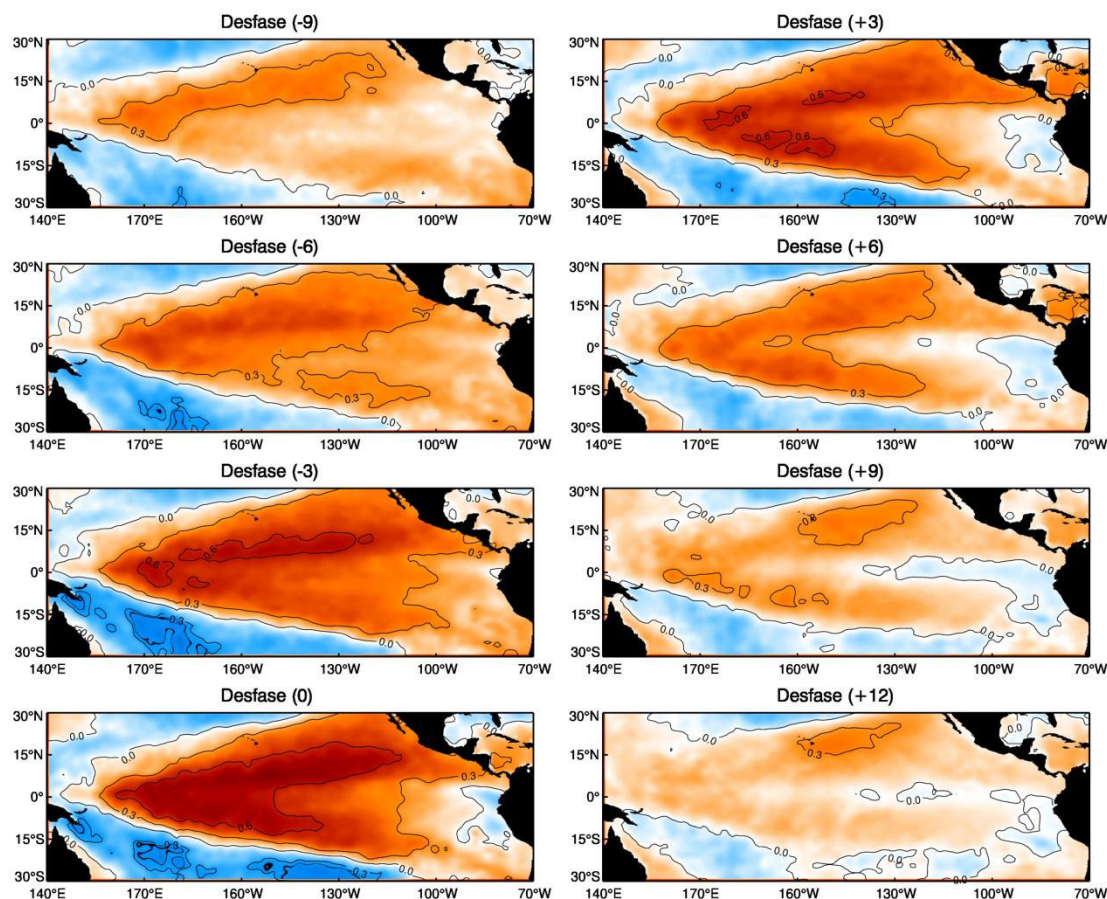


Figura 5.36: Coeficientes de correlación desfasada entre las anomalías del Pacífico Tropical y el índice CP (similar a lo presentado por kao and Yu, 2009).

En cuanto a los impactos frente a la costa peruana, durante un fuerte evento El Niño del Pacífico Este, las aguas cálidas son transportadas dentro del sistema de afloramiento, incrementando la TSM. Durante el evento El Niño del Pacífico Central, no existe tal transporte y las temperaturas son alteradas de forma débil frente al Perú (Dewitte et al, 2012). Un evento El Niño del Pacífico Este está caracterizado por intensas precipitaciones en la costa norte y centro del Perú (debido al desplazamiento de la zona de convección hacia el Este). Durante un evento El Niño del Pacífico Central las temperaturas no varían significativamente, asociándose a condiciones oceánicas más frías (no hay precipitaciones).

Al igual que lo realizado con los índices E y C, se procedió a comparar los índices EP y CP estimados (usando un algoritmo hecho en el lenguaje IDL) y los índices calculados por Kao y Yu (datos obtenidos de <https://www.ess.uci.edu/~yu/2OSC/>), obteniéndose nuevamente altas correlaciones (0.88 y 0.83) para los índices EP y CP (figuras 5.37 y 5.38), lo cual indica que los índices obtenidos en este trabajo de tesis representan de forma adecuado a los dos tipos de ENSO (ENSO del Pacífico Central y Pacífico Este).

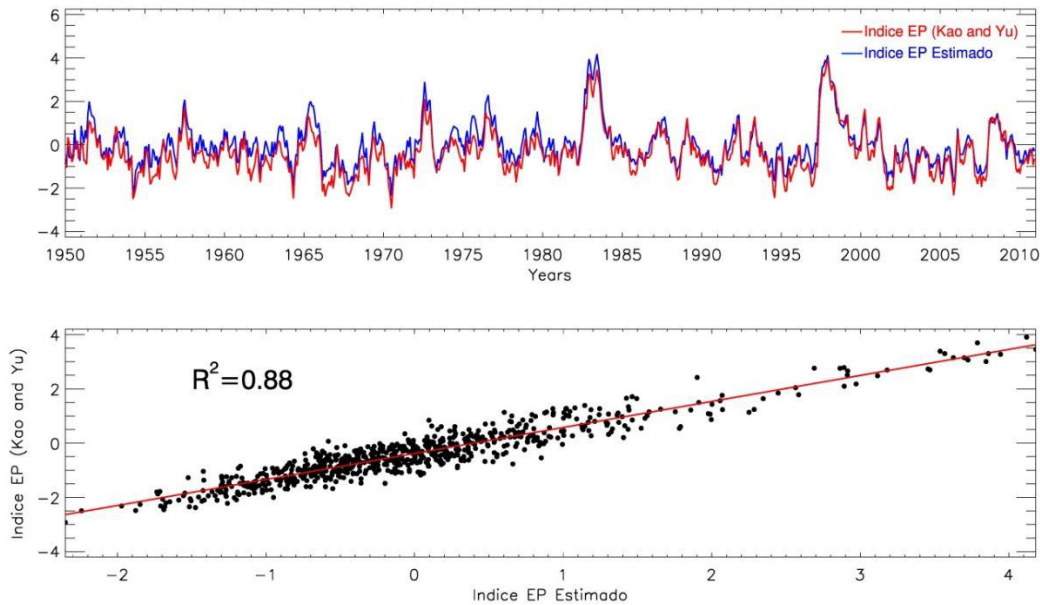


Figura 5.37: Comparación entre el índice EP estimado y el índice EP calculado por Kao y Yu, 2009.

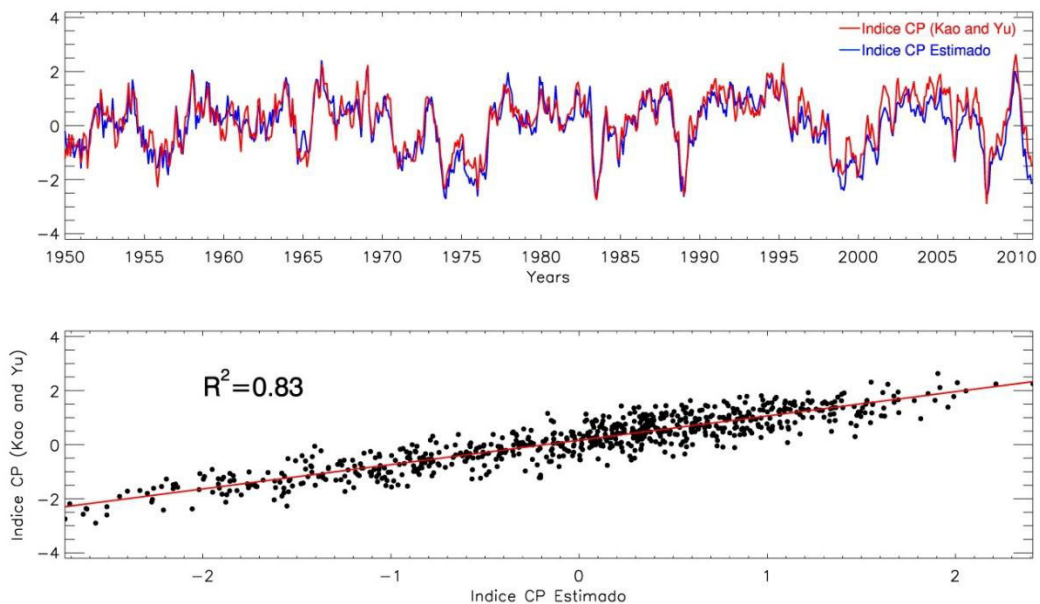


Figura 5.38: Comparación entre el índice CP estimado y el índice CP calculado por Kao y Yu, 2009.

CAPÍTULO VI

Conclusiones y Recomendaciones

- a) El análisis de las Funciones Ortogonales Empíricas (FOE) demuestra ser una técnica multivariada descriptiva que permite reducir la dimensionalidad de los datos si es que estos aportan poca información. En este sentido se ha podido reducir la información contenida en cientos de imágenes a unos pocos modos o patrones espaciales de variabilidad, los cuales dependen directamente del porcentaje de la varianza acumulada.
- b) El análisis FOE tiene amplias aplicaciones debido a que ésta revela estructuras ocultas en conjuntos de datos complejos usando soluciones analíticas del álgebra lineal.
- c) El análisis FOE aplicado a datos de TSM ha permitido identificar fenómenos globales como EL NIÑO, LA NIÑA y EL NIÑO MODOKI, los cuales están representados en los primeros dos modos de variabilidad interanual.
- d) Aunque útil, el análisis FOE no es infalible. Las FOE dependen del dominio espacial considerado, están sujetos a restricciones de ortogonalidad, y pueden no ser separables si representan porcentajes similares de la varianza.
- e) Las FOE son empíricamente determinados y por lo tanto no son necesariamente equivalentes a los modos dinámicos de un sistema. Por lo tanto siempre es prudente confirmar los patrones FOE con técnicas simples como mapas de correlación y análisis de regresión lineal.
- f) Con los resultados obtenidos en este trabajo, las FOE demuestran ser una herramienta imprescindible si se quiere realizar algún estudio de variabilidad y teleconexiones sobre una región geográfica.
- g) Los índices E y C son por construcción independientes (no correlacionados) y describen los regímenes que representan a) Eventos cálidos extremos y b) Eventos fríos ordinarios y eventos moderadamente cálidos.

- h)** La amplitud del CP El Niño tiende a ser menor que la del EP El Niño, mientras que el CP La Niña es a menudo más fuerte que el EP La Niña. La razón de esta diferencia en el comportamiento entre las fases de El Niño y La Niña de estos dos tipos de ENOS aún es tema de investigación por parte de la comunidad científica.
- i)** El hecho de que El Niño pueda ocurrir a veces en el Pacífico oriental, a veces en el Pacífico central y, a veces, simultáneamente en ambas partes del Pacífico, ha sugerido que puede existir más de un tipo de ENOS, cuyos mecanismos de generación y extremos climáticos asociados pueden ser diferente.
- j)** La diferencia en la dinámica de los dos tipos de El Niño debe documentarse para comprender la tendencia a largo plazo de la variabilidad del Pacífico tropical.

ANEXO 1: Fundamentos algebraicos de las Componentes Principales

Desafortunadamente no hay una terminología y notación universalmente aceptada para PCA. Por ejemplo algunos autores en la literatura oceanográfica utilizan los términos ‘Análisis de Factores’, cuando en realidad están realizando un análisis de componentes principales. Incluso dentro del dominio de PCA, términos tales como ‘autovectores’ y ‘funciones ortogonales empíricas’, los cuales vamos a distinguir después, son confusos o usados como sinónimos. Nuestra definición de ‘Componente Principal’ varía con respecto a algunos libros de ‘Análisis de Factores’. Nos hemos esforzado en definir una consistente notación y terminología.

Inicializaremos el enfoque del dominio del análisis de las componentes principales en una sola dirección a lo largo de un conjunto de datos bivariados. En tal dominio las operaciones algebraicas de PCA son fácilmente visualizadas y proporcionará una base para los enfoques más formales.

Comencemos por analizar un conjunto de datos bivariados. Podemos considerar por ejemplo las temperaturas promedio del aire máximas y mínimas registradas frente al puerto del Callao en el mes de enero durante el periodo 1982-2011. Por lo tanto para cada uno de los 30 eneros se obtiene un par de promedios, y estos son graficados como 30 puntos en la figura A.1

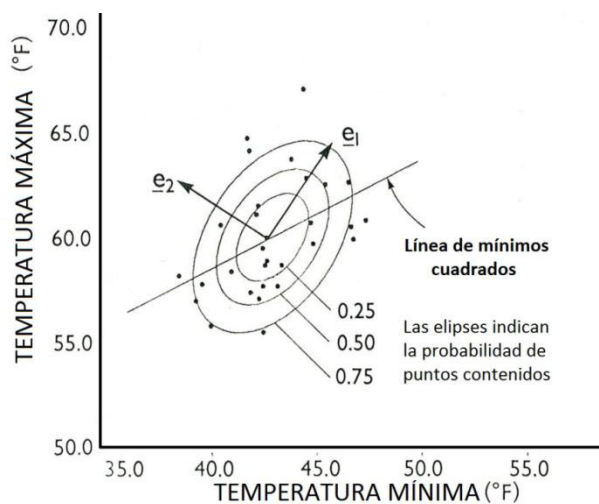


Figura A.1. Los 30 puntos representan 30 pares de temperaturas mínimas y máximas frente al puerto del Callao (1982-2011). Los autovectores e_1 y e_2 se muestran alineados a lo largo de los ejes principales de las elipses, las cuales sirven para definir la distribución de probabilidad normal bivariada del conjunto de puntos (adaptado de Preisendorfer).

Centrado y Rotación de los Datos

El análisis PCA de los datos bivariados comienza cuando investigamos la varianza de la nube de los 30 puntos de temperatura alrededor de su promedio y a lo largo de un conjunto de direcciones distintas a los ejes horizontal y vertical. Para llevar a cabo esta investigación se establece una rotación para el marco geométrico mostrado en la figura A.2

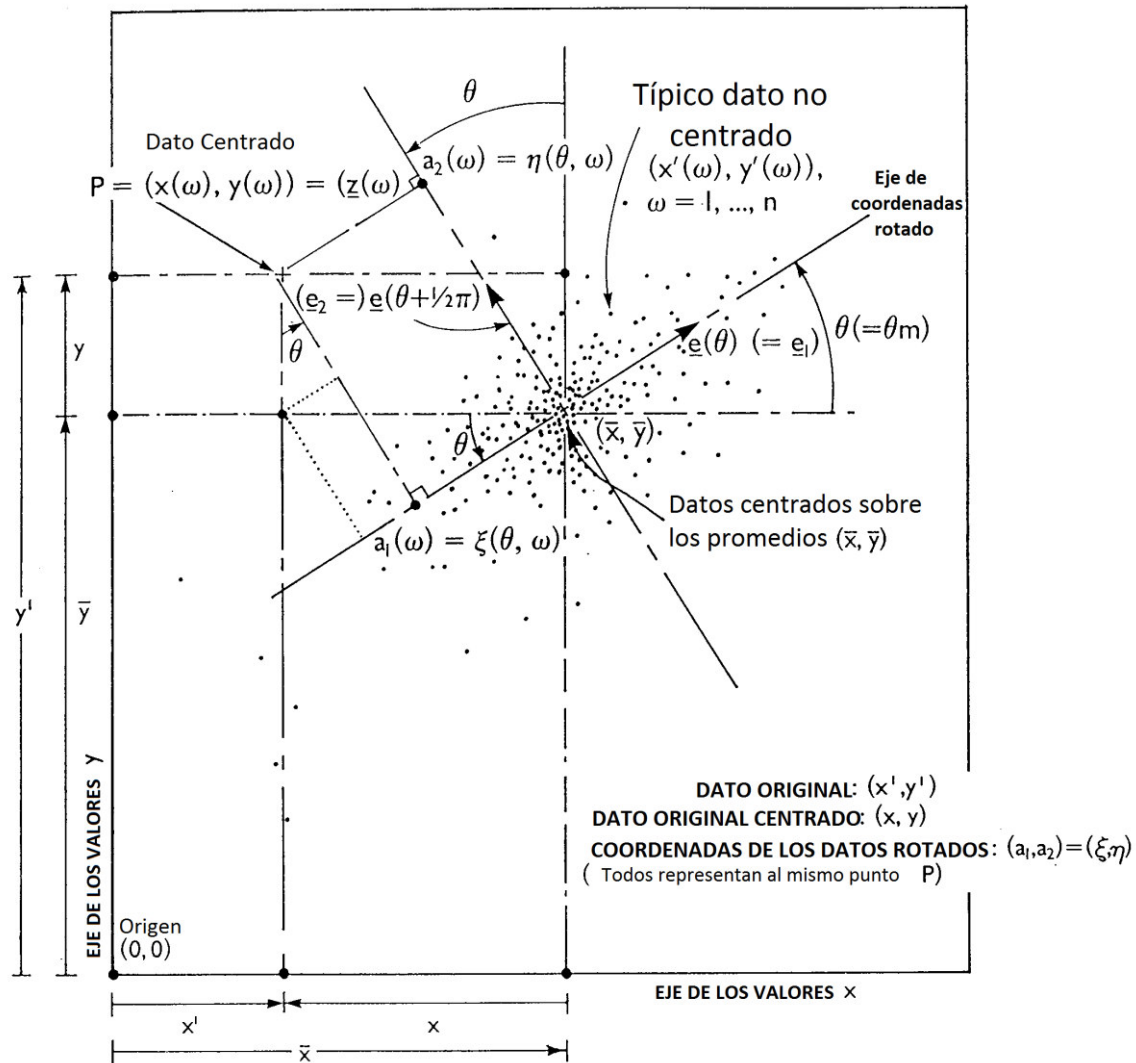


Figura A.2 Geometría del análisis de componentes principales para el caso bivariado. El diagrama puede ser visto de dos maneras: Para una rotación general θ definimos $\xi_{(\theta,\omega)}, \eta_{(\theta,\omega)}$; y para una rotación específica θ_m (primer ángulo principal) se definen los autovectores e_1 y e_2 y las componentes principales $a_1(\omega), a_2(\omega)$ de un punto general $z(\omega)$. (adaptado de Preisendorfer).

Sea $\{(\mathbf{x}'_{(\omega)}, \mathbf{y}'_{(\omega)}): \omega = 1, 2, \dots, n\}$ el conjunto de temperaturas en el conjunto de datos originales. Supongamos por ejemplo que $\mathbf{y}'_{(\omega)}$ y $\mathbf{x}'_{(\omega)}$ son respectivamente las temperaturas mínimas y máximas promedio del mes de enero para el año $\omega = 1, 2, \dots, n$. Entonces los promedios de estos valores son:

$$\bar{x} = \frac{1}{n} \sum_{\omega=1}^n \mathbf{x}'_{(\omega)} \quad y \quad \bar{y} = \frac{1}{n} \sum_{\omega=1}^n \mathbf{y}'_{(\omega)} \quad (A1)$$

Vamos a escribir

$$\mathbf{x}_{(\omega)} = \mathbf{x}'_{(\omega)} - \bar{x} \quad y \quad \mathbf{y}_{(\omega)} = \mathbf{y}'_{(\omega)} - \bar{y} \quad (A2)$$

Estos valores representan las anomalías de $\mathbf{x}'_{(\omega)}$ y $\mathbf{y}'_{(\omega)}$ sobre los 30 años. Hemos trasladado el marco del conjunto de datos a un nuevo marco de coordenadas centradas sobre el punto promedio (\bar{x}, \bar{y}) del conjunto de datos. Para considerar la varianza del conjunto de datos con respecto a su punto promedio y a lo largo de la dirección definida por θ en la figura A2, rotamos el marco xy θ radianes en sentido positivo con respecto a (\bar{x}, \bar{y}) para establecer un nuevo marco de referencia.

Un punto $P = (\mathbf{x}_{(\omega)}, \mathbf{y}_{(\omega)})$ en el marco rotado centrado en (\bar{x}, \bar{y}) tiene ahora coordenadas $(\xi_{(\theta, \omega)}, \eta_{(\theta, \omega)})$ dados por:

$$\begin{aligned} \xi_{(\theta, \omega)} &= \mathbf{x}_{(\omega)} \cos \theta + \mathbf{y}_{(\omega)} \sin \theta, \quad 0 \leq \theta \leq 2\pi \\ \eta_{(\theta, \omega)} &= -\mathbf{x}_{(\omega)} \sin \theta + \mathbf{y}_{(\omega)} \cos \theta, \quad \omega = 1, \dots, n \end{aligned} \quad (A3)$$

Varianzas en el Marco Rotado

En el marco rotado podemos calcular la varianza del conjunto de números $\{\xi_{(\theta, \omega)}: \omega = 1, 2, \dots, n\}$. Esto representará la varianza del conjunto original de datos $\{(\mathbf{x}'_{(\omega)}, \mathbf{y}'_{(\omega)})\}$. A lo largo de la dirección del eje ξ , es decir a lo largo del vector $\mathbf{e}_{\theta} = (\cos \theta, \sin \theta)^T$, donde T denota la transpuesta de la matriz.

De la ecuación (A2) tenemos:

$$\sum_{\omega=1}^n \mathbf{x}_{(\omega)} = 0 \quad ; \quad \sum_{\omega=1}^n \mathbf{y}_{(\omega)} = 0 \quad (A4a)$$

Y por (A3) tenemos:

$$\sum_{\omega=1}^n \xi_{(\theta,\omega)} = 0 \quad ; \quad \sum_{\omega=1}^n \eta_{(\theta,\omega)} = 0 \quad (A4b)$$

Demostración de (A4a):

$$\sum_{\omega=1}^n \mathbf{x}_{(\omega)} = \sum_{\omega=1}^n (\mathbf{x}'_{(\omega)} - \bar{\mathbf{x}}) = \sum_{\omega=1}^n \mathbf{x}'_{(\omega)} - \sum_{\omega=1}^n \bar{\mathbf{x}} = n\bar{\mathbf{x}} - n\bar{\mathbf{x}} = 0$$

Demostración de (A4b)

$$\sum_{\omega=1}^n \xi_{(\theta,\omega)} = \sum_{\omega=1}^n \{ \mathbf{x}_{(\omega)} \cos \theta + \mathbf{y}_{(\omega)} \sin \theta \} = \underbrace{\left\{ \sum_{\omega=1}^n \mathbf{x}_{(\omega)} \right\}}_0 \cos \theta + \underbrace{\left\{ \sum_{\omega=1}^n \mathbf{y}_{(\omega)} \right\}}_0 \sin \theta = 0$$

La varianza $S^2(\theta)$ del conjunto de datos a lo largo de \mathbf{e}_θ utilizando (A3) es:

$$\begin{aligned} S^2(\theta) &= (n-1)^{-1} \sum_{\omega=1}^n \xi_{(\theta,\omega)}^2 \\ &= (n-1)^{-1} \sum_{\omega=1}^n (\mathbf{x}_{(\omega)} \cos \theta + \mathbf{y}_{(\omega)} \sin \theta)^2 \end{aligned} \quad (A5)$$

$$S^2(\theta) = S_{xx} \cos^2 \theta + 2S_{xy} \sin \theta \cos \theta + S_{yy} \sin^2 \theta \quad (A6)$$

Donde:

$$\left. \begin{aligned} S_{xx} &= (n-1)^{-1} \sum_{\omega=1}^n \mathbf{x}_{(\omega)}^2 \\ S_{yy} &= (n-1)^{-1} \sum_{\omega=1}^n \mathbf{y}_{(\omega)}^2 \\ S_{xy} &= (n-1)^{-1} \sum_{\omega=1}^n \mathbf{x}_{(\omega)} \mathbf{y}_{(\omega)} \end{aligned} \right\} \quad (A7)$$

Demostración de (A6):

$$S^2(\theta) = (n-1)^{-1} \sum_{\omega=1}^n \xi_{(\theta,\omega)}^2$$

$$\begin{aligned}
S^2(\theta) &= (n-1)^{-1} \sum_{\omega=1}^n (\mathbf{x}_{(\omega)} \cos \theta + \mathbf{y}_{(\omega)} \sin \theta)^2 \\
S^2(\theta) &= (n-1)^{-1} \sum_{\omega=1}^n (\mathbf{x}_{(\omega)}^2 \cos^2 \theta + 2\mathbf{x}_{(\omega)}\mathbf{y}_{(\omega)} \sin \theta \cos \theta + \mathbf{y}_{(\omega)}^2 \sin^2 \theta) \\
S^2(\theta) &= \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n \mathbf{x}_{(\omega)}^2 \right\}}_{S_{xx}} \cos^2 \theta + \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n \mathbf{y}_{(\omega)}^2 \right\}}_{S_{yy}} \sin^2 \theta \\
&\quad + 2 \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n \mathbf{x}_{(\omega)}\mathbf{y}_{(\omega)} \right\}}_{S_{xy}} \sin \theta \cos \theta
\end{aligned}$$

$$S^2(\theta) = S_{xx} \cos^2 \theta + 2S_{xy} \sin \theta \cos \theta + S_{yy} \sin^2 \theta$$

S_{xx} y S_{yy} son respectivamente las varianzas de los datos originales de temperaturas máximas y mínimas. De (A6) vemos que cuando $\theta = 0$, entonces $S^2(0) = S_{xx}$; y cuando $\theta = \frac{\pi}{2}$, entonces $S^2\left(\frac{\pi}{2}\right) = S_{yy}$. Como θ varía de 0 a $\frac{\pi}{2}$, $S^2(\theta)$ toma todos los valores entre S_{xx} y S_{yy} . S_{xy} es la covarianza del conjunto de datos.

Ángulos Principales

Los ángulos principales del conjunto de datos bivariados $\{(\mathbf{x}'_{(\omega)}, \mathbf{y}'_{(\omega)}) : \omega = 1, 2, \dots, n\}$ ocurren en los valores máximos y mínimos de $S^2(\theta)$ en (A6). Estos ángulos pueden ser encontrados haciendo:

$$\frac{d(S^2(\theta))}{d\theta} = (S_{yy} - S_{xx}) \sin 2\theta + 2S_{xy} \cos 2\theta = 0 \quad (A8)$$

Demostración de (A8)

$$\frac{d(S^2(\theta))}{d\theta} = \frac{d}{d\theta} (S_{xx} \cos^2 \theta + 2S_{xy} \sin \theta \cos \theta + S_{yy} \sin^2 \theta) = 0$$

$$\frac{d(S^2(\theta))}{d\theta} = -S_{xx} \underbrace{(2 \sin \theta \cos \theta)}_{\sin 2\theta} + S_{yy} \underbrace{(2 \sin \theta \cos \theta)}_{\sin 2\theta} + 2S_{xy} \underbrace{(\cos^2 \theta - \sin^2 \theta)}_{\cos 2\theta} = 0$$

$$\frac{d(S^2(\theta))}{d\theta} = (S_{yy} - S_{xx})\sin 2\theta + 2S_{xy} \cos 2\theta = 0$$

Los valores requeridos de θ serán por lo tanto dados por las soluciones θ' de la ecuación trascendental. De (A8) obtenemos:

$$\tan 2\theta' = \frac{2S_{xy}}{S_{xx} - S_{yy}} \quad (A9)$$

Observe que si θ' satisface (A9) también lo hace $\theta' + n\pi/2$, donde 'n' es cualquier entero, por lo tanto hay un número infinito de θ' que satisfacen (A9).

Generalmente existen dos ángulos definibles de forma única θ' y $\theta' + \pi/2$, en los cuales ocurren los valores extremos $S^2(\theta)$. Uno de estos valores es un máximo y el otro un mínimo, como puede verse al ver los signos relativos de:

$$\left. \frac{d^2(S^2(\theta))}{d\theta^2} \right|_{\theta=\theta'} = -4S_{xy} / \sin 2\theta' \quad \text{para } \theta' \text{ y } \theta' + \frac{\pi}{2}$$

Demostración:

$$\frac{d^2(S^2(\theta))}{d\theta^2} = \frac{d}{d\theta} \left(\frac{d(S^2(\theta))}{d\theta} \right) = \frac{d}{d\theta} ((S_{yy} - S_{xx})\sin 2\theta + 2S_{xy} \cos 2\theta)$$

$$\frac{d^2(S^2(\theta))}{d\theta^2} = 2 \underbrace{(S_{yy} - S_{xx})}_{(A9)} \cos 2\theta - 4S_{xy} \sin 2\theta$$

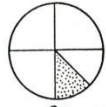








$$\frac{d^2(S^2(\theta))}{d\theta^2} = 2 \left(\frac{-2S_{xy}}{\tan 2\theta} \right) \cos 2\theta - 4S_{xy} \sin 2\theta$$

$$\frac{d^2(S^2(\theta))}{d\theta^2} = -4S_{xy} \frac{\cos^2 2\theta}{\sin 2\theta} - 4S_{xy} \sin 2\theta = -4S_{xy} \left(\frac{\cos^2 2\theta + \sin^2 2\theta}{\sin 2\theta} \right) = \frac{-4S_{xy}}{\sin 2\theta}$$

$$\Rightarrow \left. \frac{d^2(S^2(\theta))}{d\theta^2} \right|_{\theta=\theta'} = \frac{-4S_{xy}}{\sin 2\theta'}$$

El ángulo θ_m a lo largo del cual $S^2(\theta)$ es un máximo es el primer ángulo principal θ_m . El valor de θ_m puede ser fijado de forma única en todos los casos al adoptar las convenciones mostradas en la figura A.3. Primero adoptamos la rama principal de la relación $\frac{1}{2} \arctan$ (que está en el intervalo $(-\pi/4, \pi/4)$) como se muestra en la figura. Una vez que θ' es

encontrada entonces se establece θ_m , dependiendo de los signos y valores de S_{yy} y $(S_{xx} - S_{yy})$ en (A9).

	$s_{xy} < 0$	$s_{xy} = 0$	$s_{xy} > 0$
$s_{xx} - s_{yy} < 0$	 $\theta_m = \theta' - \pi/2$	 $\theta_m = \pi/2$	 $\theta_m = \theta' + \pi/2$
$s_{xx} - s_{yy} = 0$	 $\theta_m = -\pi/4$	 $\theta_m = 0$	 $\theta_m = \pi/4$
$s_{xx} - s_{yy} > 0$	 $\theta_m = \theta'$	 $\theta_m = 0$	 $\theta_m = \theta'$

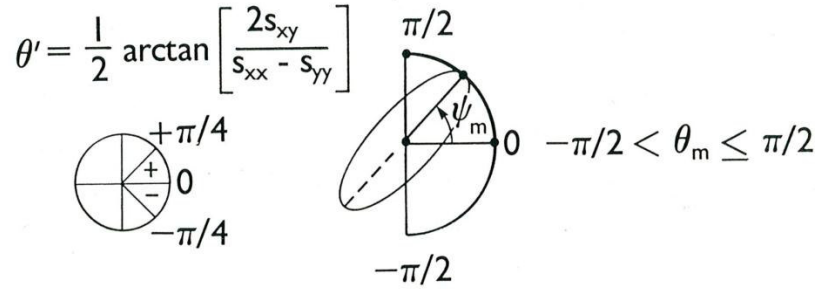


Figura A.3. Convenciones usadas para determinar el ángulo principal θ_m de un conjunto de datos bivariados

Hay 9 distintas posibilidades para estos signos y valores. Por ejemplo si $(S_{xx} - S_{yy}) < 0$ y $S_{xy} > 0$ entonces la convención requiere que $\theta_m = \theta' + \pi/2$, donde θ' es dado por la rama principal de la relación $\frac{1}{2} \arctan$ evaluada en $\frac{2S_{xy}}{(S_{xx} - S_{yy})} < 0$. Por lo tanto en este ejemplo θ' está en alguna parte entre $(-\frac{\pi}{4}, 0)$ y por tanto θ_m en alguna parte entre $(\frac{\pi}{4}, \frac{\pi}{2})$, la cual se muestra como una región sombreada en el círculo. En general θ_m se encuentra en alguna parte en el intervalo $(-\frac{\pi}{2}, \frac{\pi}{2})$ para cada uno de los 9 casos graficados en la figura A3.

Varianzas Principales

Las dos varianzas principales del conjunto de datos ocurren en los ángulos principales y son:

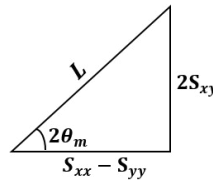
$$\mathbf{S}_{11} = S^2(\theta_m) \quad \text{y} \quad \mathbf{S}_{22} = S^2(\theta_m + \pi/2) \quad (\text{A10})$$

La ecuación (A9) puede ser escrita en una forma equivalente:

$$\cos 2\theta_m = \pm \frac{(S_{xx} - S_{yy})}{\left[(S_{xx} - S_{yy})^2 + 4S_{xy}^2\right]^{1/2}} \quad (\text{A11})$$

Demostración de (A11):

De la ecuación (A9): $\tan 2\theta_m = \frac{2S_{xy}}{S_{xx} - S_{yy}} \Rightarrow$



$$\Rightarrow L^2 = 4S_{xy}^2 + (S_{xx} - S_{yy})^2$$

$$\Rightarrow \cos 2\theta_m = \frac{(S_{xx} - S_{yy})}{L} = \pm \frac{(S_{xx} - S_{yy})}{\left[4S_{xy}^2 + (S_{xx} - S_{yy})^2\right]^{1/2}}$$

Usando (A6) para θ_m y $\theta_m + \pi/2$ encontramos la siguiente ecuación:

$$\left. \begin{matrix} \mathbf{S}_{11} \\ \mathbf{S}_{22} \end{matrix} \right\} = \frac{1}{2} \left[(S_{xx} - S_{yy}) \pm \left[(S_{xx} - S_{yy})^2 + 4S_{xy}^2 \right]^{1/2} \right] \quad (\text{A12})$$

Demostración de (A12)

Recordar que: $\sin(\theta_m + \pi/2) = \cos \theta_m$

$$\cos(\theta_m + \pi/2) = -\sin \theta_m$$

De la ecuación (A6) para θ_m y $\theta_m + \pi/2$

$$\mathbf{S}_{11} = S^2(\theta_m) = S_{xx} \cos^2 \theta_m + 2S_{xy} \sin \theta_m \cos \theta_m + S_{yy} \sin^2 \theta_m \dots \dots \dots (\alpha)$$

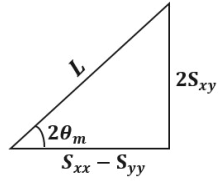
$$\mathbf{S}_{22} = S^2(\theta_m + \pi/2)$$

$$= S_{xx} \sin^2 \theta_m - 2S_{xy} \sin \theta_m \cos \theta_m + S_{yy} \cos^2 \theta_m \dots \dots \dots (\beta)$$

\Rightarrow De (α) y (β) :

$$S_{11} + S_{22} = S_{xx} + S_{yy} \dots \dots \dots (\delta)$$

$$S_{11} - S_{22} = 2S_{xy} \underbrace{\sin 2\theta_m} + (S_{xx} - S_{yy}) \underbrace{\cos 2\theta_m}$$



$$\Rightarrow L^2 = 4S_{xy}^2 + (S_{xx} - S_{yy})^2$$

$$S_{11} - S_{22} = 2S_{xy} \left(\frac{2S_{xy}}{L} \right) + (S_{xx} - S_{yy}) \left(\frac{S_{xx} - S_{yy}}{L} \right)$$

$$S_{11} - S_{22} = \frac{4S_{xy}^2}{\left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2}} + \frac{(S_{xx} - S_{yy})^2}{\left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2}}$$

$$S_{11} - S_{22} = \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \dots\dots\dots (\varphi)$$

\Rightarrow De (δ) y (φ) :

$$\begin{aligned} (\delta) + (\varphi) &\Rightarrow 2S_{11} = S_{xx} + S_{yy} + \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \\ &\Rightarrow S_{11} = \frac{1}{2} \left[(S_{xx} + S_{yy}) + \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \right] \end{aligned}$$

$$\begin{aligned} (\delta) - (\varphi) &\Rightarrow 2S_{22} = S_{xx} + S_{yy} - \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \\ &\Rightarrow S_{22} = \frac{1}{2} \left[(S_{xx} + S_{yy}) - \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \right] \end{aligned}$$

$$\therefore \begin{Bmatrix} S_{11} \\ S_{22} \end{Bmatrix} = \frac{1}{2} \left[(S_{xx} + S_{yy}) \pm \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \right]$$

En general para cualquier θ y cualquier conjunto de datos bivariados tenemos que:

$$0 \leq \mathbf{S}_{22} \leq S^2(\theta) \leq \mathbf{S}_{11} \quad (A13)$$

Y en particular si $S_{xx} \leq S_{yy}$ (como es el caso del ejemplo) entonces

$$0 \leq \mathbf{S}_{22} \leq S_{xx} \leq S_{yy} \leq \mathbf{S}_{11} \quad (A14)$$

Covarianza Principal

Asociada con las varianzas principales \mathbf{S}_{11} y \mathbf{S}_{22} hay una covarianza principal $S_{12}(\theta_m)$. En general para cualquier θ definimos la covarianza $S_{12}(\theta)$ de un conjunto de datos en el marco rotado un ángulo θ como:

$$\mathbf{S}_{12}(\boldsymbol{\theta}) = (n-1)^{-1} \sum_{\omega=1}^n \boldsymbol{\xi}_{(\theta,\omega)} \boldsymbol{\eta}_{(\theta,\omega)} \quad (A15)$$

El cual utilizando la ecuación (A3) se reduce a:

$$\mathbf{S}_{12}(\boldsymbol{\theta}) = \frac{1}{2}(S_{yy} - S_{xx}) \text{sen}2\theta + S_{xy} \cos2\theta \quad (A16)$$

Demostración de (A16)

$$\begin{aligned} S_{12}(\theta) &= (n-1)^{-1} \sum_{\omega=1}^n \xi_{(\theta,\omega)} \eta_{(\theta,\omega)} \\ S_{12}(\theta) &= (n-1)^{-1} \sum_{\omega=1}^n (x_{(\omega)} \cos\theta + y_{(\omega)} \text{sen}\theta) (-x_{(\omega)} \text{sen}\theta + y_{(\omega)} \cos\theta) \\ S_{12}(\theta) &= (n-1)^{-1} \sum_{\omega=1}^n (-x_{(\omega)}^2 \text{sen}\theta \cos\theta - x_{(\omega)} y_{(\omega)} \text{sen}^2\theta \\ &\quad + x_{(\omega)} y_{(\omega)} \cos^2\theta + y_{(\omega)}^2 \text{sen}\theta \cos\theta) \\ S_{12}(\theta) &= \left\{ \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n y_{(\omega)}^2 \right\}}_{S_{yy}} - \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n x_{(\omega)}^2 \right\}}_{S_{xx}} \right\} \text{sen}\theta \cos\theta \\ &\quad + \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n x_{(\omega)} y_{(\omega)} \right\}}_{S_{xy}} (\cos^2\theta - \text{sen}^2\theta) \\ S_{12}(\theta) &= (S_{yy} - S_{xx}) \underbrace{\text{sen}\theta \cos\theta}_{\left(\frac{\text{sen}2\theta}{2}\right)} + S_{xy} \underbrace{(\cos^2\theta - \text{sen}^2\theta)}_{\cos2\theta} \\ S_{12}(\theta) &= \frac{1}{2}(S_{yy} - S_{xx}) \text{sen}2\theta + S_{xy} \cos2\theta \end{aligned}$$

Para el caso donde $\theta = \theta_m$, usando la ecuación (A9) con $\theta' = \theta_m$ encontramos:

$$\mathbf{S}_{12}(\boldsymbol{\theta}_m) = 0 \quad (A17)$$

Demostración de (A17)

$$S_{12}(\boldsymbol{\theta}_m) = \frac{1}{2} \underbrace{(S_{yy} - S_{xx})}_{(A9)} \text{sen}2\boldsymbol{\theta}_m + S_{xy}\text{cos}2\boldsymbol{\theta}_m$$

$$S_{12}(\boldsymbol{\theta}_m) = \frac{1}{2} \left(\frac{-2S_{xy}}{\text{tg}2\boldsymbol{\theta}_m} \right) \text{sen}2\boldsymbol{\theta}_m + S_{xy}\text{cos}2\boldsymbol{\theta}_m$$

$$S_{12}(\boldsymbol{\theta}_m) = -S_{xy}\text{cos}2\boldsymbol{\theta}_m + S_{xy}\text{cos}2\boldsymbol{\theta}_m$$

$$S_{12}(\boldsymbol{\theta}_m) = 0$$

Por lo tanto la covarianza del conjunto de datos es cero en el marco principal (sistema de ejes principales). En otras palabras los dos conjuntos de datos $\{\boldsymbol{\xi}_{(\theta,\omega)}: \omega = 1, 2 \dots \dots, n\}$, $\{\boldsymbol{\eta}_{(\theta,\omega)}: \omega = 1, 2 \dots \dots, n\}$ no están correlacionados en el sistema de ejes principales.

Direcciones Principales

Las direcciones principales (o vectores) de un conjunto de datos bivariados son obtenidas al usar $\boldsymbol{\theta}_m$ y $\boldsymbol{\theta}_m + \pi/2$ en $\mathbf{e}_{1(\theta)} = [\text{cos}\theta, \text{sen}\theta]^T$. Por lo tanto las direcciones principales son dos vectores en el plano de datos.

$$\mathbf{e}_1 = \mathbf{e}_{(\theta_m)} = [\text{cos}\theta, \text{sen}\theta]^T \quad (A18)$$

$$\mathbf{e}_2 = \mathbf{e}_{(\theta_m+\pi/2)} = [-\text{sen}\theta, \text{cos}\theta]^T$$

Claramente estos vectores tienen longitud unitaria y son ortogonales.

$$\mathbf{e}_1^T \mathbf{e}_1 = \mathbf{e}_2^T \mathbf{e}_2 = \mathbf{1} \quad (A19)$$

$$\mathbf{e}_1^T \mathbf{e}_2 = \mathbf{e}_2^T \mathbf{e}_1 = \mathbf{0}$$

Más adelante se demostrará que \mathbf{e}_1 y \mathbf{e}_2 son las “Funciones Ortogonales Empíricas” (FOE’s) o los autovectores del conjunto de datos bivariados.

Componentes Principales: Direcciones principales como vectores base

Los vectores principales \mathbf{e}_1 y \mathbf{e}_2 en (A18) forman una base para el espacio de datos bidimensionales de las figuras A.1 y A.2 en el siguiente sentido. Escribimos un punto $(\mathbf{x}_{(\omega)}, \mathbf{y}_{(\omega)})$ como un vector (una matriz 2x1 de los datos centrados):

$$\mathbf{z}_{(\omega)} = (\mathbf{x}_{(\omega)}, \mathbf{y}_{(\omega)})^T \quad (A20)$$

Las componentes $\mathbf{a}_{1(\omega)}$ y $\mathbf{a}_{2(\omega)}$ de $\mathbf{z}_{(\omega)}$ a lo largo de las direcciones principales \mathbf{e}_1 y \mathbf{e}_2 son dados por:

$$\mathbf{a}_{1(\omega)} = \mathbf{z}_{(\omega)}^T \mathbf{e}_1 \quad ; \omega = 1, \dots, n \quad (A21)$$

$$\mathbf{a}_{2(\omega)} = \mathbf{z}_{(\omega)}^T \mathbf{e}_2$$

Estas son las componentes principales de $\mathbf{z}_{(\omega)}$. El vector $\mathbf{z}_{(\omega)}$, para cada “ ω ”, puede ser representado como una combinación lineal de \mathbf{e}_1 y \mathbf{e}_2 .

$$\mathbf{z}_{(\omega)} = \mathbf{a}_{1(\omega)} \mathbf{e}_1 + \mathbf{a}_{2(\omega)} \mathbf{e}_2 \quad ; \omega = 1, \dots, n \quad (A22)$$

Esta es la representación de los datos deseada, mostrando que \mathbf{e}_1 y \mathbf{e}_2 forman una base del espacio de datos. Anticipando algunos aspectos dinámicos del análisis PCA a ser considerado después, podemos también pensar de las $\mathbf{a}_{j(\omega)}$ como las amplitudes del conjunto de datos, con \mathbf{e}_1 y \mathbf{e}_2 siendo los modos de variación de $\mathbf{z}_{(\omega)}$. La forma escalar de (A22) es encontrada, escribiendo primero (A22) como:

$$\begin{bmatrix} \mathbf{x}_{(\omega)} \\ \mathbf{y}_{(\omega)} \end{bmatrix} = \mathbf{a}_{1(\omega)} \begin{bmatrix} \cos\theta_m \\ \sin\theta_m \end{bmatrix} + \mathbf{a}_{2(\omega)} \begin{bmatrix} -\sin\theta_m \\ \cos\theta_m \end{bmatrix}$$

Y luego:

$$\mathbf{x}_{(\omega)} = \mathbf{a}_{1(\omega)} \cos\theta_m - \mathbf{a}_{2(\omega)} \sin\theta_m \quad (A23)$$

$$\mathbf{y}_{(\omega)} = \mathbf{a}_{1(\omega)} \sin\theta_m + \mathbf{a}_{2(\omega)} \cos\theta_m$$

El conjunto de ecuaciones (A23) es precisamente la forma de la solución de (A3) para $\mathbf{x}_{(\omega)}$ y $\mathbf{y}_{(\omega)}$ en términos de $\xi_{(\theta, \omega)}$ y $\eta_{(\theta, \omega)}$. Haciendo $\theta = \theta_m$ en (A3) y comparando con (A23) encontramos que:

$$\mathbf{a}_{1(\omega)} = \xi_{(\theta, \omega)} \quad ; \omega = 1, \dots, n \quad (A24)$$

$$\mathbf{a}_{2(\omega)} = \eta_{(\theta, \omega)}$$

Demostración de (A24)

En (A3) hacemos $\theta = \theta_m$ y usamos (A23)

$$\xi_{(\theta_m, \omega)} = \underbrace{\mathbf{x}_{(\omega)}}_{(A23)} \cos \theta_m + \underbrace{\mathbf{y}_{(\omega)}}_{(A23)} \sen \theta_m$$

$$\xi_{(\theta_m, \omega)} = (\mathbf{a}_{1(\omega)} \cos \theta_m - \mathbf{a}_{2(\omega)} \sen \theta_m) \cos \theta_m + (\mathbf{a}_{1(\omega)} \sen \theta_m + \mathbf{a}_{2(\omega)} \cos \theta_m) \sen \theta_m$$

$$\xi_{(\theta_m, \omega)} = \mathbf{a}_{1(\omega)} \cos^2 \theta_m + \mathbf{a}_{1(\omega)} \sen^2 \theta_m$$

$$\xi_{(\theta_m, \omega)} = \mathbf{a}_{1(\omega)}$$

De forma análoga se encuentra que:

$$\mathbf{a}_{2(\omega)} = \boldsymbol{\eta}_{(\theta, \omega)}$$

De (A21) y (A23) concluimos que las componentes principales $\mathbf{a}_{1(\omega)}$ y $\mathbf{a}_{2(\omega)}$ de un vector de datos $\mathbf{z}_{(\omega)} = [\mathbf{x}_{(\omega)}, \mathbf{y}_{(\omega)}]^T$ son las proyecciones de $\mathbf{z}_{(\omega)}$ sobre las direcciones principales \mathbf{e}_1 y \mathbf{e}_2 . Además los valores de los datos son exactamente reproducibles usando las direcciones principales y las componentes principales del conjunto de datos bivariados. En este sentido \mathbf{e}_1 y \mathbf{e}_2 forman una base del espacio de datos.

Representación Matricial

De la ecuación (A22) tenemos que:

$$\mathbf{z}_{(\omega)} = [\mathbf{e}_1 \quad \mathbf{e}_2] \begin{bmatrix} \mathbf{a}_{1(\omega)} \\ \mathbf{a}_{2(\omega)} \end{bmatrix}$$

Calculando la transpuesta:

$$\mathbf{z}_{(\omega)}^T = [\mathbf{a}_{1(\omega)} \quad \mathbf{a}_{2(\omega)}] \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{bmatrix}$$

$$\mathbf{z}_{(\omega)}^T = \mathbf{a}_{1(\omega)} \mathbf{e}_1^T + \mathbf{a}_{2(\omega)} \mathbf{e}_2^T$$

Que en forma matricial puede ser escrita como:

$$\begin{bmatrix} \mathbf{z}_{(1)}^T \\ \mathbf{z}_{(2)}^T \\ \vdots \\ \mathbf{z}_{(n)}^T \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{1(1)} \\ \mathbf{a}_{1(2)} \\ \vdots \\ \mathbf{a}_{1(n)} \end{bmatrix} \mathbf{e}_1^T + \begin{bmatrix} \mathbf{a}_{2(1)} \\ \mathbf{a}_{2(2)} \\ \vdots \\ \mathbf{a}_{2(n)} \end{bmatrix} \mathbf{e}_2^T \quad (A25)$$

Definiendo nuevas matrices:

$$\underline{\mathbf{Z}} = \begin{bmatrix} \mathbf{z}^T_{(1)} \\ \mathbf{z}^T_{(2)} \\ \vdots \\ \mathbf{z}^T_{(n)} \end{bmatrix} \quad (n \times 2) \quad (\text{A26})$$

$$\underline{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_{1(1)} & \mathbf{a}_{2(1)} \\ \mathbf{a}_{1(2)} & \mathbf{a}_{2(2)} \\ \vdots & \vdots \\ \mathbf{a}_{1(n)} & \mathbf{a}_{2(n)} \end{bmatrix} \quad (n \times 2) \quad (\text{A27})$$

$$\underline{\mathbf{E}} = [\mathbf{e}_1 \quad \mathbf{e}_2] \quad (2 \times 2) \quad (\text{A28})$$

El conjunto de ecuaciones (A22) se convierte en la representación matricial del conjunto de datos:

$$\underline{\mathbf{Z}} = \underline{\mathbf{A}} \underline{\mathbf{E}}^T \quad (\text{A29})$$

El conjunto de componentes principales en (A21) se convierte en la “matriz de amplitudes”

$$\underline{\mathbf{A}} = \underline{\mathbf{Z}} \underline{\mathbf{E}} \quad (\text{A30})$$

Y las relaciones de ortonormalidad (A19) son expresadas como:

$$\underline{\mathbf{E}}^T \underline{\mathbf{E}} = \underline{\mathbf{E}} \underline{\mathbf{E}}^T = \underline{\mathbf{I}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (\text{A31})$$

Propiedad de las Componentes Principales

Usando la ecuación (A27) establecemos la más importante propiedad de las componentes principales

$$\begin{aligned} \underline{\mathbf{A}}^T \underline{\mathbf{A}} &= \begin{bmatrix} \mathbf{a}_{1(1)} & \mathbf{a}_{2(1)} \\ \vdots & \vdots \\ \mathbf{a}_{1(n)} & \mathbf{a}_{2(n)} \end{bmatrix}^T \begin{bmatrix} \mathbf{a}_{1(1)} & \mathbf{a}_{2(1)} \\ \vdots & \vdots \\ \mathbf{a}_{1(n)} & \mathbf{a}_{2(n)} \end{bmatrix} \\ \underline{\mathbf{A}}^T \underline{\mathbf{A}} &= \begin{bmatrix} \sum_{\omega=1}^n \mathbf{a}_{1(\omega)}^2 & \sum_{\omega=1}^n \mathbf{a}_{1(\omega)} \mathbf{a}_{2(\omega)} \\ \sum_{\omega=1}^n \mathbf{a}_{1(\omega)} \mathbf{a}_{2(\omega)} & \sum_{\omega=1}^n \mathbf{a}_{2(\omega)}^2 \end{bmatrix} \end{aligned} \quad (\text{A32})$$

$$\underline{\mathbf{A}}^T \underline{\mathbf{A}} = (n-1) \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix} \quad (2 \times 2) \quad (\text{A33})$$

De las ecuaciones (A15 y (A17) se deduce que los elementos fuera de la diagonal de $\underline{\mathbf{A}}^T \underline{\mathbf{A}}$ son cero. Los elementos de la diagonal principal toman su forma utilizando las ecuaciones (A5), (A10) y (A24), estableciendo $\theta = \theta_m$ y $\theta = \theta_m + \pi/A$. Así el conjunto de componentes principales $\{\mathbf{a}_{1(\omega)}: \omega = 1, \dots, n\}$ y $\{\mathbf{a}_{2(\omega)}: \omega = 1, \dots, n\}$ no están correlacionadas y tienen varianzas S_{11} y S_{22} respectivamente.

Invariancia de la Varianza Total en la Rotación

Existen algunas conexiones útiles entre las varianzas del conjunto de datos originales S_{xx} y S_{yy} y sus varianzas principales S_{11} y S_{22} . Elevando al cuadrado cada lado de las ecuaciones en (A23), sumando cada ecuación sobre ω y usando la no correlación entre $\mathbf{a}_{1(\omega)}$ y $\mathbf{a}_{2(\omega)}$ encontramos usando la definición en (A7) que:

$$\begin{aligned} S_{xx} &= S_{11} \cos^2 \theta_m + S_{22} \sin^2 \theta_m \\ S_{yy} &= S_{11} \sin^2 \theta_m + S_{22} \cos^2 \theta_m \end{aligned} \quad (A34)$$

Demostración de (A34)

De la ecuación (A23) tenemos:

$$\begin{aligned} \mathbf{x}_{(\omega)} &= \mathbf{a}_{1(\omega)} \cos \theta_m - \mathbf{a}_{2(\omega)} \sin \theta_m \\ \Rightarrow \mathbf{x}_{(\omega)}^2 &= \mathbf{a}_{1(\omega)}^2 \cos^2 \theta_m + \mathbf{a}_{2(\omega)}^2 \sin^2 \theta_m - \mathbf{a}_{1(\omega)} \mathbf{a}_{2(\omega)} \sin 2\theta_m \end{aligned}$$

Sumando cada ecuación sobre ω y multiplicándolos por $(n-1)^{-1}$

$$\begin{aligned} \underbrace{(n-1)^{-1} \sum_{\omega=1}^n \mathbf{x}_{(\omega)}^2}_{S_{xx}} &= \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n \mathbf{a}_{1(\omega)}^2 \right\}}_{S_{11}} \cos^2 \theta_m + \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n \mathbf{a}_{2(\omega)}^2 \right\}}_{S_{22}} \sin^2 \theta_m \\ &\quad - \underbrace{\left\{ (n-1)^{-1} \sum_{\omega=1}^n \mathbf{a}_{1(\omega)} \mathbf{a}_{2(\omega)} \right\}}_0 \sin 2\theta_m \end{aligned}$$

$$\Rightarrow S_{xx} = S_{11} \cos^2 \theta_m + S_{22} \sin^2 \theta_m$$

De forma análoga se obtiene:

$$S_{yy} = S_{11}\sin^2\theta_m + S_{22}\cos^2\theta_m$$

Sumando ambas ecuaciones encontramos que:

$$S_{xx} + S_{yy} = S_{11} + S_{22} \quad (A35)$$

La ecuación (A35) muestra que la varianza total del conjunto de datos bivariados es igual a la suma de las varianzas principales, y la varianza total es por lo tanto invariante después de rotar las coordenadas hacia el sistema de ejes principales.

Restando las ecuaciones (A34) encontramos:

$$S_{xx} - S_{yy} = (S_{11} - S_{22})\cos 2\theta_m \quad (A36)$$

De la ecuación (A9) haciendo $\theta' = \theta_m$

$$\tan 2\theta_m = \frac{2S_{xy}}{S_{xx} - S_{yy}}$$

Y usando la ecuación (A36)

$$\tan 2\theta_m = \frac{2S_{xy}}{(S_{11} - S_{22})\cos 2\theta_m} \quad (A37)$$

Obtenemos:

$$S_{xy} = \frac{1}{2}(S_{11} - S_{22})\sin 2\theta_m \quad (A38)$$

Esto sugiere que si las varianzas principales de un conjunto de datos bivariados son iguales, entonces la covarianza de los datos originales es cero (covarianza nula).

Note también que cuando $S_{11} = S_{22}$, entonces por (A36) y (A38) $S_{xx} = S_{yy}$ y $S_{xy} = 0$.

Si se aplica la condición anterior a la ecuación (A6) notamos que $S^2(\theta)$ es independiente de θ . Entonces los vectores \mathbf{e}_1 y \mathbf{e}_2 en (A18) pueden tomar cualquier orientación mientras aún mantengan su ortogonalidad. Nuestra convención en figura A3 establece $\theta_m = 0$ en este caso.

Lo siguiente en observar es que el producto $S_{11}S_{22}$ usando (A12)

$$\begin{Bmatrix} S_{11} \\ S_{22} \end{Bmatrix} = \frac{1}{2} \left[(S_{xx} + S_{yy}) \pm \left[4S_{xy}^2 + (S_{xx} - S_{yy})^2 \right]^{1/2} \right]$$

$$\begin{aligned}\Rightarrow S_{11}S_{22} &= \frac{1}{4} \left[(S_{xx} + S_{yy})^2 - [4S_{xy}^2 + (S_{xx} - S_{yy})^2] \right] \\ \Rightarrow S_{11}S_{22} &= \frac{1}{4} [4S_{xx}S_{yy} - 4S_{xy}^2]\end{aligned}$$

Puede reducirse a:

$$S_{11}S_{22} = (S_{xx}S_{yy} - S_{xy}^2) \geq 0 \quad (A39)$$

Alternativamente si definimos una cantidad adimensional

$$r_{xy} = \frac{S_{xy}}{S_{xx}^{1/2} S_{yy}^{1/2}} \quad (A40)$$

Reemplazando (A40) en (A39)

$$S_{11}S_{22} = S_{xx}S_{yy}(1 - r_{xy}^2) \geq 0 \quad (A41)$$

La cantidad adimensional r_{xy} debe satisfacer $-1 \leq r_{xy} \leq 1$, y es el coeficiente de correlación de un conjunto de datos bivariados. En general por (A41) si $r_{xy} = 0$ entonces $S_{11}S_{22} = S_{xx}S_{yy}$. Sin embargo si $r_{xy} = 1$ entonces $S_{22} = 0$ (pero no necesariamente $S_{11} = 0$). De (A38) podemos concluir que si $S_{11} = S_{22}$ entonces $r_{xy} = 0$.

Varianzas principales para conjuntos de datos estandarizados

Consideremos las formas de las varianzas de un conjunto de datos bivariados estandarizados. Un conjunto de datos ha sido estandarizado cuando toman la forma:

$$\mathbf{u}_{(\omega)} = \frac{\mathbf{x}'_{(\omega)} - \bar{\mathbf{x}}}{S_{xx}^{1/2}} \quad y \quad \mathbf{v}_{(\omega)} = \frac{\mathbf{y}'_{(\omega)} - \bar{\mathbf{y}}}{S_{yy}^{1/2}}$$

para $\omega = 1, \dots, n$

Es obvio que \mathbf{u} y \mathbf{v} son adimensionales y que las varianzas muestrales S_{uu} y S_{vv} son de valor unitario. Sin embargo para un conjunto de datos estandarizados la covarianza muestral S_{uv} es numéricamente igual al coeficiente de correlación muestral r_{uv} .

Por (A36) y (A38) tenemos que

$$S_{uu} - S_{vv} = (S_{11} - S_{22})\cos 2\theta_m$$

$$S_{uv} = \frac{1}{2}(S_{11} - S_{22})\text{sen}2\theta_m$$

Si $S_{11} \neq S_{22}$ el primer ángulo principal θ_m para el conjunto de datos estandarizados es $\pi/4$ o $-\pi/4$ dependiendo de si $S_{uv} > 0$ o $S_{uv} < 0$ respectivamente. Si utilizamos ahora la ecuación (A12):

$$\begin{Bmatrix} S_{11} \\ S_{22} \end{Bmatrix} = \frac{1}{2} [(S_{uu} + S_{vv}) \pm [(S_{uu} - S_{vv})^2 + 4S_{uv}^2]^{1/2}]$$

Reemplazando $S_{uu} = S_{vv} = 1$ y $S_{uv} = \mathbf{r}_{uv}$ obtenemos:

$$\begin{Bmatrix} S_{11} \\ S_{22} \end{Bmatrix} = 1 \pm \mathbf{r}_{uv} \quad (A42)$$

PCA y estimación de los parámetros estadísticos de poblaciones normales

Si el conjunto de datos bivariados $\{(\mathbf{x}'_{(\omega)}, \mathbf{y}'_{(\omega)}): \omega = 1, 2, \dots, n\}$ es una muestra aleatoria de una población normal, entonces el análisis de componentes principales de la muestra produce estimados de los parámetros de la población. Sea a_1 y a_2 las variables aleatorias independientes obtenidas de dos poblaciones normales univariados con funciones de densidad de probabilidad

$$\mathbf{p}_1(a_1) = (2\pi\sigma_{11})^{-1/2} \exp\{-a_1^2/(2\sigma_{11})\} \quad (A43)$$

$$\mathbf{p}_2(a_2) = (2\pi\sigma_{22})^{-1/2} \exp\{-a_2^2/(2\sigma_{22})\} \quad (A44)$$

Por tanto a_j esta distribuida normalmente con media cero y varianza σ_{jj} . En símbolos $a_j \sim N(0, \sigma_{jj})$ con $j = 1, 2$. La función densidad de probabilidad conjunta $\mathbf{p}_{12}(a_1, a_2)$ para estas variables independientes es entonces:

$$\begin{aligned} \mathbf{p}_{12}(a_1, a_2) &= \mathbf{p}_1(a_1) \mathbf{p}_2(a_2) \\ \mathbf{p}_{12}(a_1, a_2) &= \frac{1}{2\pi(\sigma_{11}\sigma_{22})^{1/2}} \exp\left\{-\frac{1}{2}\left(\frac{a_1^2}{\sigma_{11}} + \frac{a_2^2}{\sigma_{22}}\right)\right\} \end{aligned} \quad (A45)$$

Si retornamos a la configuración bivariada general y extraemos una muestra aleatoria $\{(\mathbf{x}'_{(\omega)}, \mathbf{y}'_{(\omega)}): \omega = 1, 2, \dots, n\}$ de una población normal bivariada y luego se procesa

aquella muestra, encontramos que el conjunto de componentes principales $a_{1(\omega)}$ y $a_{2(\omega)}$ son variables aleatorias normales no correlacionadas. Estas variables normales son independientes y sus probabilidades son gobernadas por las ecuaciones (A43) y (A44), donde ahora las varianzas principales S_{11} y S_{22} de la muestra estiman las varianzas poblacionales σ_{11} y σ_{22} . Por lo tanto la distribución de las variables normales en el marco de varianza principal de la figura A2 es estimada por (A45) con S_{11} y S_{22} reemplazando a σ_{11} y σ_{22} . El retorno a la forma usual $p(x, y)$ de $p_{12}(a_1, a_2)$ en el marco de coordenadas originales puede ser realizado usando (A24) en (A45), recordando (A23) y reemplazando σ_{11} y σ_{22} por S_{11} y S_{22} . La simplificación de la forma algebraica resultante es lograda al usar (A38) y (A41). El resultado es:

$$p(x, y) = \frac{1}{2\pi (S_{xx}S_{yy}(1 - r_{xy}^2))^{\frac{1}{2}}} \exp \left\{ \frac{-1}{2(1 - r_{xy}^2)} \left(\frac{x^2}{S_{xx}} - \frac{2xy r_{xy}}{(S_{xx}S_{yy})^{\frac{1}{2}}} + \frac{y^2}{S_{yy}} \right) \right\} \quad (A46)$$

Donde y y x son los datos centrados dados en (A2), S_{xx} y S_{yy} son definidos en (A7) y r_{xy} es definido en (A40). Algunas elipses de $p(x, y)$ son mostradas en la figura A1.

PCA y la construcción de experimentos de Monte Carlo

Las observaciones anteriores sugieren cómo el método PCA puede ser usado para construir una población normal bivariada para propósitos de Monte Carlo. Suponga que se requiere muestrear una población normal bivariada con varianzas σ_{xx} y σ_{yy} y covarianza $\sigma_{xy} = \rho_{xy}\sigma_{xx}^{1/2}\sigma_{yy}^{1/2}$ donde ρ_{xy} es el coeficiente de correlación de la población. Por (A12) usando σ en lugar de S , podemos encontrar las varianzas principales σ_{11} y σ_{22} de la población.

$$\left. \begin{matrix} \sigma_{11} \\ \sigma_{22} \end{matrix} \right\} = \frac{1}{2} \left[(\sigma_{xx} - \sigma_{yy}) \pm \left[(\sigma_{xx} - \sigma_{yy})^2 + 4\sigma_{xy}^2 \right]^{1/2} \right] \quad (A47)$$

“Si σ_{xx} , σ_{yy} son obtenidos de una configuración física real, entonces S_{22} será no negativa. Sin embargo, si en un experimento de datos simulados uno escoge arbitrariamente σ_{xx} , σ_{yy} y σ_{xy} entonces (A12) puede retornar un S_{22} negativo, indicando que las elecciones no fueron realistas.”

Autovalores y autovectores de las matrices de Covarianza y Dispersión

Se mostrará que las direcciones principales \underline{e}_1 y \underline{e}_2 en (A18) son los autovectores, y también que las varianzas principales S_{11} y S_{22} en (A12) son los autovalores de la matriz de covarianza del conjunto de datos.

$$\underline{C} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \quad (A48)$$

Para demostrar esto debemos establecer las ecuaciones

$$\underline{C} \underline{e}_j = S_{jj} \underline{e}_j \quad ; j = 1, 2 \quad (A49)$$

Por ejemplo si $j = 1$ entonces:

$$\begin{aligned} \underline{C} \underline{e}_1 &= \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \begin{bmatrix} \cos\theta_m \\ \sin\theta_m \end{bmatrix} \\ &= \begin{bmatrix} S_{xx} \cos\theta_m + S_{xy} \sin\theta_m \\ S_{xy} \cos\theta_m + S_{yy} \sin\theta_m \end{bmatrix} \end{aligned}$$

Usando las ecuaciones (A34) y (A38)

$$\begin{aligned} &= \begin{bmatrix} (S_{11} \cos^2\theta_m + S_{22} \sin^2\theta_m) \cos\theta_m + \left(\frac{1}{2}(S_{11} - S_{22}) \sin 2\theta_m\right) \sin\theta_m \\ \left(\frac{1}{2}(S_{11} - S_{22}) \sin 2\theta_m\right) \cos\theta_m + (S_{11} \sin^2\theta_m + S_{22} \cos^2\theta_m) \sin\theta_m \end{bmatrix} \\ &= \begin{bmatrix} (S_{11} \cos^2\theta_m + S_{22} \sin^2\theta_m) \cos\theta_m + (S_{11} - S_{22}) \sin^2\theta_m \cos\theta_m \\ (S_{11} - S_{22}) \cos^2\theta_m \sin\theta_m + (S_{11} \sin^2\theta_m + S_{22} \cos^2\theta_m) \sin\theta_m \end{bmatrix} \\ &= \begin{bmatrix} S_{11} \cos\theta_m \\ S_{11} \sin\theta_m \end{bmatrix} \\ &= S_{11} \begin{bmatrix} \cos\theta_m \\ \sin\theta_m \end{bmatrix} \\ &= S_{jj} \underline{e}_j \end{aligned} \quad (A50)$$

El caso de $j = 2$ puede ser establecido similarmente. Observe que si $S_{11} = S_{22}$, entonces $S_{xx} = S_{yy}$ y $S_{xy} = 0$. En este caso es claro en el desarrollo de (A50) que 2 vectores ortogonales cualesquiera \underline{e}_1 y \underline{e}_2 pueden ser autovectores de \underline{C} , y por lo tanto la unicidad de \underline{e}_1 y \underline{e}_2 no es posible. Para propósitos de simplicidad presentamos la ecuación (A49) en una forma ligeramente diferente. Note que de acuerdo a (A33) las medidas naturales alternas de la dispersión de un conjunto de datos es $(n-1)S_{11}$ y $(n-1)S_{22}$, en lugar de S_{11} y S_{22} . En la misma forma podemos usar $(n-1)S_{xx}$, $(n-1)S_{yy}$ y $(n-1)S_{xy}$ en lugar de S_{xx} , S_{yy} y S_{xy} .

Definamos la siguiente matriz como la matriz de dispersión de un conjunto de datos bivariados.

$$\underline{\mathbf{S}} = \begin{bmatrix} \sum_{\omega=1}^n \mathbf{x}^2_{(\omega)} & \sum_{\omega=1}^n \mathbf{x}_{(\omega)}\mathbf{y}_{(\omega)} \\ \sum_{\omega=1}^n \mathbf{x}_{(\omega)}\mathbf{y}_{(\omega)} & \sum_{\omega=1}^n \mathbf{y}^2_{(\omega)} \end{bmatrix} \quad (A51)$$

Estableciendo:

$$l_j = (n - 1)S_{jj} \quad ; j = 1,2 \quad (A52)$$

La matriz de autovalores de $\underline{\mathbf{S}}$ es:

$$\underline{\mathbf{L}} = \begin{bmatrix} l_1 & 0 \\ 0 & l_2 \end{bmatrix} \quad (A53)$$

La ecuación (A49) para $\underline{\mathbf{e}}_1$ y $\underline{\mathbf{e}}_2$ y S_{11} y S_{22} puede ser escrita como:

$$\underline{\mathbf{S}}\underline{\mathbf{e}}_j = l_j \underline{\mathbf{e}}_j \quad ; j = 1,2 \quad (A54)$$

En forma matricial (recordando A28) estas ecuaciones toman la siguiente forma:

$$\underline{\mathbf{S}}\underline{\mathbf{E}} = \underline{\mathbf{E}}\underline{\mathbf{L}} \quad (A55)$$

Demostración de (A55)

$$\begin{aligned} \underline{\mathbf{S}}\underline{\mathbf{E}} &= \begin{bmatrix} \sum_{\omega=1}^n \mathbf{x}^2_{(\omega)} & \sum_{\omega=1}^n \mathbf{x}_{(\omega)}\mathbf{y}_{(\omega)} \\ \sum_{\omega=1}^n \mathbf{x}_{(\omega)}\mathbf{y}_{(\omega)} & \sum_{\omega=1}^n \mathbf{y}^2_{(\omega)} \end{bmatrix} \begin{bmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{bmatrix} \\ \underline{\mathbf{S}}\underline{\mathbf{E}} &= \begin{bmatrix} (n-1)S_{xx} & (n-1)S_{xy} \\ (n-1)S_{xy} & (n-1)S_{yy} \end{bmatrix} \begin{bmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{bmatrix} \\ \underline{\mathbf{S}}\underline{\mathbf{E}} &= (n-1) \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \begin{bmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{bmatrix} \\ \underline{\mathbf{S}}\underline{\mathbf{E}} &= (n-1) \begin{bmatrix} S_{xx}\cos\theta_m + S_{xy}\sin\theta_m & -S_{xx}\sin\theta_m + S_{xy}\cos\theta_m \\ S_{xy}\cos\theta_m + S_{yy}\sin\theta_m & -S_{xy}\sin\theta_m + S_{yy}\cos\theta_m \end{bmatrix} \end{aligned}$$

Aplicando (A34) y (A38)

$$\underline{\mathbf{S}}\underline{\mathbf{E}} = (n-1) \begin{bmatrix} S_{11}\cos\theta_m & -S_{22}\sin\theta_m \\ S_{11}\sin\theta_m & S_{22}\cos\theta_m \end{bmatrix}$$

$$\underline{\mathbf{S}} \underline{\mathbf{E}} = \underbrace{\begin{bmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{bmatrix}}_{\underline{\mathbf{E}}} \underbrace{\begin{bmatrix} (n-1)S_{11} & 0 \\ 0 & (n-1)S_{22} \end{bmatrix}}_{\underline{\mathbf{L}}}$$

$$\therefore \underline{\mathbf{S}} \underline{\mathbf{E}} = \underline{\mathbf{E}} \underline{\mathbf{L}}$$

La ecuación (A55) es nuestro puente a la teoría general del análisis de componentes principales. La conversión de (A54) a (A49) es realizada dividiendo $\underline{\mathbf{S}}$ y l_j por $(n-1)$. Observe que los autovectores en (A49) y (A54) son el mismo par $\underline{\mathbf{e}}_1$, $\underline{\mathbf{e}}_2$ y que estos han sido únicamente determinados por la convención que la primera componente de $\underline{\mathbf{e}}_1$ es no negativa.

Análisis de Componentes Principales: Campos escalares de valor real

Consideremos el análisis de componentes principales (PCA) de un conjunto de datos de valor real $\underline{\mathbf{Z}}'$ en la forma de una matriz de dimensiones $n \times p$.

$$\underline{\mathbf{Z}}' = \begin{bmatrix} \mathbf{z}'_{(1,1)} & \dots & \mathbf{z}'_{(1,p)} \\ \vdots & \ddots & \vdots \\ \mathbf{z}'_{(n,1)} & \dots & \mathbf{z}'_{(n,p)} \end{bmatrix} \quad (\text{A56})$$

Para fijar ideas se sugiere que el elemento escalar real $\mathbf{z}'_{(t,x)}$ ($t = 1, 2, \dots, n$ y $x = 1, 2, \dots, p$) de $\underline{\mathbf{Z}}'$ sea visto como el valor de la temperatura superficial del mar en el punto (t, x) , es decir en el tiempo t y posición x .

Centrado del conjunto de datos en el tiempo

El primer paso en el análisis PCA de $\underline{\mathbf{Z}}'$ es centrar los valores de $\mathbf{z}'_{(t,x)}$ en sus promedios sobre el índice t . Así escribimos para cada $x = 1, 2, \dots, p$.

$$\bar{z}_{(x)} = n^{-1} \sum_{t=1}^n \mathbf{z}'_{(t,x)} \quad (\text{A57})$$

Y a continuación se forman las anomalías

$$\mathbf{z}_{(t,x)} = \mathbf{z}'_{(t,x)} - \bar{z}_{(x)} \quad (\text{A58})$$

Usando estos valores centrados en el tiempo $\mathbf{z}_{(t,x)}$ formamos una nueva matriz $\underline{\mathbf{Z}}$ tal como la matriz (A56). Para muchas de las operaciones resulta sumamente útil pensar en cada fila

de $\underline{\mathbf{Z}}$ como puntos en el espacio euclidiano p -dimensional " E_p ". Entonces para cada valor de $t = 1, 2, \dots, n$.

$$\underline{\mathbf{z}}_{(t, \cdot)} = \underline{\mathbf{z}}_{(t)} = [z_{(t,1)}, \dots, z_{(t,p)}]^T \quad (p \times 1) \quad (A59)$$

Entonces

$$\underline{\mathbf{Z}} = [\underline{\mathbf{z}}_{(1)}, \dots, \underline{\mathbf{z}}_{(n)}]^T \quad (n \times p) \quad (A60)$$

Dispersión de la muestra y Matriz de dispersión

Definimos ahora la dispersión del conjunto de datos $\underline{\mathbf{Z}}$, considerado como un conjunto de n puntos en E_p a lo largo de un eje con vector unitario $\underline{\mathbf{e}} = [\mathbf{e}_{(1)}, \dots, \mathbf{e}_{(p)}]^T$. La proyección de $\underline{\mathbf{z}}_{(t)}$ sobre $\underline{\mathbf{e}}$ es:

$$\underline{\mathbf{z}}_{(t)}^T \underline{\mathbf{e}} = \underline{\mathbf{e}}^T \underline{\mathbf{z}}_{(t)} = \sum_{x=1}^p z_{(t,x)} e_{(x)}$$

Donde: $\underline{\mathbf{e}}^T \underline{\mathbf{e}} = 1$

Definimos la dispersión $\psi_{(\underline{\mathbf{e}})}$ de $\underline{\mathbf{Z}}$ a lo largo de $\underline{\mathbf{e}}$ como:

$$\begin{aligned} \psi_{(\underline{\mathbf{e}})} &= \sum_{x=1}^p [\underline{\mathbf{z}}_{(t)}^T \underline{\mathbf{e}}]^2 \\ &= \sum_{x=1}^p [\underline{\mathbf{e}}^T \underline{\mathbf{z}}_{(t)}] [\underline{\mathbf{z}}_{(t)}^T \underline{\mathbf{e}}] \\ \psi_{(\underline{\mathbf{e}})} &= \underline{\mathbf{e}}^T \left[\sum_{t=1}^n \underline{\mathbf{z}}_{(t)} \underline{\mathbf{z}}_{(t)}^T \right] \underline{\mathbf{e}} \geq 0 \end{aligned} \quad (A61)$$

Podemos pensar de $\psi_{(\underline{\mathbf{e}})}$ como una dispersión (o varianza) muestral: Para cada elección de $\underline{\mathbf{e}}$ tendremos en " E_p " una dispersión de $\underline{\mathbf{Z}}$ a lo largo de $\underline{\mathbf{e}}$. En analogía con $\underline{\mathbf{S}}$ en (A51) definimos la matriz de dispersión $(p \times p)$ $\underline{\mathbf{S}}$ de $\underline{\mathbf{Z}}$.

$$\underline{\mathbf{S}} = \sum_{t=1}^n \underline{\mathbf{z}}_{(t)} \underline{\mathbf{z}}_{(t)}^T = \underline{\mathbf{Z}}^T \underline{\mathbf{Z}} \quad (A62)$$

Si $s_{(x,x')}$ es el elemento de $\underline{\mathbf{S}}$ en su x -ésima fila y x' -ésima columna, entonces por (A62) tenemos que:

$$\mathbf{s}_{(x,x')} = \sum_{t=1}^n \mathbf{z}_{(t,x)} \mathbf{z}_{(t,x')}^T, \quad x, x' = 1, \dots, p \quad (A63)$$

Podemos ahora escribir la dispersión $\psi_{(\underline{\mathbf{e}})}$ en (A61) de $\underline{\mathbf{Z}}$ a lo largo de $\underline{\mathbf{e}}$ como:

$$\psi_{(\underline{\mathbf{e}})} = \underline{\mathbf{e}}^T \underline{\mathbf{S}} \underline{\mathbf{e}} \geq 0 \quad (A64)$$

Se observa de (A62) o (A63) que es $\underline{\mathbf{S}}$ simétrica, es decir:

$$\underline{\mathbf{S}}^T = \underline{\mathbf{S}} \quad (A65)$$

A partir de esta propiedad se obtiene una relación muy importante. Si $\vec{\mathbf{a}}$ y $\vec{\mathbf{b}}$ son dos vectores de dimensiones $(p \times 1)$ entonces:

$$\vec{\mathbf{a}}^T \underline{\mathbf{S}} \vec{\mathbf{b}} = \vec{\mathbf{b}}^T \underline{\mathbf{S}} \vec{\mathbf{a}} \quad (A66)$$

Las autoestructuras del Análisis de Componentes Principales

Consideremos ahora el siguiente problema que consiste en encontrar las direcciones $\underline{\mathbf{e}}$ en E_p para los cuales $\psi_{(\underline{\mathbf{e}})}$ tiene valores extremos o estacionarios (máximos o mínimos locales). Esto nos conducirá a la introducción de los conceptos algebraicos centrales del PCA, denominados como las autoestructuras de la matriz de dispersión $\underline{\mathbf{S}}$. Si $\underline{\mathbf{e}}$ es una dirección donde $\psi_{(\underline{\mathbf{e}})}$ tiene un valor extremo, entonces para cualquier pequeña perturbación $\delta \underline{\mathbf{e}}$ de la dirección de $\underline{\mathbf{e}}$ tendríamos:

$$\psi_{(\underline{\mathbf{e}} + \delta \underline{\mathbf{e}})} = \psi_{(\underline{\mathbf{e}})} \quad (A67)$$

Ahora

$$\begin{aligned} \psi_{(\underline{\mathbf{e}} + \delta \underline{\mathbf{e}})} &= (\underline{\mathbf{e}} + \delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} (\underline{\mathbf{e}} + \delta \underline{\mathbf{e}}) \\ &= \underline{\mathbf{e}}^T \underline{\mathbf{S}} \underline{\mathbf{e}} + (\delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} \underline{\mathbf{e}} + \underbrace{\underline{\mathbf{e}}^T \underline{\mathbf{S}} (\delta \underline{\mathbf{e}})}_{(2.66)} + \\ &\quad \underbrace{(\delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} (\delta \underline{\mathbf{e}})}_0 \end{aligned}$$

Luego

$$\psi_{(\underline{\mathbf{e}} + \delta \underline{\mathbf{e}})} = \underline{\mathbf{e}}^T \underline{\mathbf{S}} \underline{\mathbf{e}} + 2(\delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} \underline{\mathbf{e}} = \psi_{(\underline{\mathbf{e}})} = \underline{\mathbf{e}}^T \underline{\mathbf{S}} \underline{\mathbf{e}}$$

Donde se ha usado la ecuación (A66) y el término $(\delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} \delta \underline{\mathbf{e}}$ se ha suprimido debido a que es un término insignificante ($\delta \underline{\mathbf{e}}$ es una perturbación muy pequeña). La ecuación (A67) implica que:

$$(\delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} \underline{\mathbf{e}} = \mathbf{0} \quad (\text{A68})$$

Sin embargo no podemos utilizar cualquier perturbación $\delta \underline{\mathbf{e}}$ de $\underline{\mathbf{e}}$, sino únicamente aquellas en las que la longitud de $\underline{\mathbf{e}} + \delta \underline{\mathbf{e}}$ permanezca unitaria (satisfaga la condición de ortogonalidad).

$$(\underline{\mathbf{e}} + \delta \underline{\mathbf{e}})^T (\underline{\mathbf{e}} + \delta \underline{\mathbf{e}}) = 1 \quad (\text{A69})$$

$$\Rightarrow (\underline{\mathbf{e}} + \delta \underline{\mathbf{e}})^T (\underline{\mathbf{e}} + \delta \underline{\mathbf{e}}) = \underbrace{\underline{\mathbf{e}}^T \underline{\mathbf{e}}}_1 + (\delta \underline{\mathbf{e}})^T \underline{\mathbf{e}} + \underline{\mathbf{e}}^T (\delta \underline{\mathbf{e}}) + \underbrace{(\delta \underline{\mathbf{e}})^T (\delta \underline{\mathbf{e}})}_0 = 1$$

$$\Rightarrow (\delta \underline{\mathbf{e}})^T \underline{\mathbf{e}} + \underline{\mathbf{e}}^T (\delta \underline{\mathbf{e}}) = 2(\delta \underline{\mathbf{e}})^T \underline{\mathbf{e}} = \mathbf{0}$$

Condición que requiere que:

$$(\delta \underline{\mathbf{e}})^T \underline{\mathbf{e}} = \mathbf{0} \quad (\text{A70})$$

Por tanto las perturbaciones $\delta \underline{\mathbf{e}}$ son ortogonales a $\underline{\mathbf{e}}$; es decir sólo cambian la dirección de $\underline{\mathbf{e}}$. Note que por convención las componentes de $\underline{\mathbf{e}}$ son adimensionales. Entonces si ' l ' es cualquier número real con las dimensiones de los elementos de $\underline{\mathbf{S}}$ podemos combinar (A68) y (A70) de tal forma que podemos escribir una identidad de la forma:

$$(\delta \underline{\mathbf{e}})^T \underline{\mathbf{S}} \underline{\mathbf{e}} - l(\delta \underline{\mathbf{e}})^T \underline{\mathbf{e}} = \mathbf{0} \quad (\text{A71})$$

De modo que

$$(\delta \underline{\mathbf{e}})^T (\underline{\mathbf{S}} \underline{\mathbf{e}} - l \underline{\mathbf{e}}) = \mathbf{0} \quad (\text{A72})$$

Para que la ecuación (A72) se cumpla es necesario y suficiente que:

$$\underline{\mathbf{S}} \underline{\mathbf{e}} = l \underline{\mathbf{e}} \quad (\text{A73})$$

Esta es la ecuación que gobierna las direcciones $\underline{\mathbf{e}}$ de dispersión extremal. Las soluciones no triviales (es decir $\underline{\mathbf{e}} \neq \mathbf{0}$) de este conjunto de p ecuaciones algebraicas lineales para las componentes de $\underline{\mathbf{e}} = [\mathbf{e}_{(1)}, \dots, \mathbf{e}_{(p)}]^T$ ocurren solo para los valores especiales de l . En la teoría del álgebra lineal se demuestra que una matriz simétrica tal como $\underline{\mathbf{S}}$ en (A73) generalmente tienen p autovectores $\underline{\mathbf{e}}_j = [\mathbf{e}_{j(1)}, \dots, \mathbf{e}_{j(p)}]^T$ en E_p (soluciones de (A73)) y p autovalores asociados reales no negativos l_j tal que:

$$\underline{\mathbf{S}} \underline{\mathbf{e}}_j = l_j \underline{\mathbf{e}}_j \quad ; (j = 1, \dots, p) \quad (\text{A74a})$$

El cual en la forma de componentes se escribe como:

$$\sum_{x'=1}^p s_{(x,x')} e_{j(x')} = l_j e_{j(x)} \quad ; (x = 1, \dots, p) \quad (A74b)$$

Esta es la generalización apropiada de (A54). De (A63) y (A74) se puede demostrar que los l_j son no negativos.

$$\underline{\mathbf{E}} = [\underline{\mathbf{e}}_1, \dots \dots \underline{\mathbf{e}}_p] \quad (p \times p) \quad (A75)$$

$$\underline{\mathbf{L}} = \text{diag}[l_1, \dots \dots l_p] \quad (p \times p) \quad (A76)$$

Donde por construcción:

$$l_1 \geq l_2 \geq l_3 \geq \dots \dots \geq l_p$$

La correspondiente generalización de (A55) es

$$\underline{\mathbf{S}} \underline{\mathbf{E}} = \underline{\mathbf{E}} \underline{\mathbf{L}} \quad (A77)$$

Observe que los autovalores l_j' para la matriz de covarianza $\underline{\mathbf{C}} = (n-1)^{-1} \underline{\mathbf{S}}$ son $l_j' = (n-1)^{-1} l_j = (n-1)^{-1} (n-1) S_{jj} = S_{jj}$. Cuando sólo el tamaño relativo de los l_j importa, uno puede normalizar los autovalores: $l_j'' = l_j / l$, $l = l_1 + l_2 + \dots \dots + l_p$, de modo que

$l_1'' + l_2'' + \dots \dots + l_p'' = 1$. Otras transformaciones de los l_j son también posibles. La matriz de dispersión es escogida para mantener todas estas opciones abiertas y el álgebra simple.

El ordenamiento de los autovalores l_j en forma decreciente sirve para identificar el correspondiente $\underline{\mathbf{e}}_j$ cuando los l_j son distintos. Cuando dos o más autovalores son iguales (el caso de autovalores múltiples) los correspondientes autovectores no son únicos.

El número de autovalores no nulos ($l_j \neq 0$) es igual al rango de la matriz $\underline{\mathbf{Z}}$. El rango de $\underline{\mathbf{Z}}$ en (A60) es el número de filas o columnas linealmente independientes en dicha matriz, lo que es equivalente a decir que es el orden de la submatriz cuadrada más grande al particionar $\underline{\mathbf{Z}}$, cuya determinante no sea nula. En general el rango de la matriz $\underline{\mathbf{Z}}$ es el menor valor entre la longitud de las series y el número de series (rango ($\underline{\mathbf{Z}}$) = min(n-1, p)). Los autovectores $\underline{\mathbf{e}}_j$ y $\underline{\mathbf{e}}_k$ pertenecientes a distintos autovalores l_j y l_k de $\underline{\mathbf{S}}$ en (A74) son ortogonales. Consideremos por la ecuación (A74):

$$\underline{\mathbf{S}} \underline{\mathbf{e}}_j = l_j \underline{\mathbf{e}}_j \quad y \quad \underline{\mathbf{S}} \underline{\mathbf{e}}_k = l_k \underline{\mathbf{e}}_k$$

Entonces:

$$\underline{\mathbf{e}}_k^T \underline{\mathbf{S}} \underline{\mathbf{e}}_j = \underline{\mathbf{e}}_k^T l_j \underline{\mathbf{e}}_j \quad (*)$$

$$\underline{\mathbf{e}}_j^T \underline{\mathbf{S}} \underline{\mathbf{e}}_k = \underline{\mathbf{e}}_j^T l_k \underline{\mathbf{e}}_k \quad (**)$$

Aplicando la ecuación (A66) en (**) obtenemos:

$$\underline{\mathbf{e}}_k^T \underline{\mathbf{S}} \underline{\mathbf{e}}_j = \underline{\mathbf{e}}_k^T l_k \underline{\mathbf{e}}_j \quad (***)$$

Restando (*) y (***)

$$0 = \underline{\mathbf{e}}_k^T \underline{\mathbf{e}}_j (l_j - l_k)$$

Por lo que $\underline{\mathbf{e}}_k^T \underline{\mathbf{e}}_j = 0$ si $l_j \neq l_k$. Si $l_j = l_k$ es posible en la práctica encontrar valores de $\underline{\mathbf{e}}_j$ y $\underline{\mathbf{e}}_k$ que satisfagan (A74) y que sean ortogonales (pero no necesariamente únicos (degeneración)). Por lo tanto en general podemos siempre encontrar autovectores que satisfagan el hecho que sean ortonormales.

$$\underline{\mathbf{e}}_j^T \underline{\mathbf{e}}_k = \delta_{jk} \quad (A78)$$

Para $(j, k = 1, 2, \dots, p)$ donde δ_{jk} es el delta de kronecker. Si nosotros definimos las componentes de $\underline{\mathbf{e}}_j$ en la forma:

$$\underline{\mathbf{e}}_j = [\mathbf{e}_{j(1)}, \dots, \mathbf{e}_{j(p)}]^T \quad (A79)$$

Entonces la versión escalar de A78 (propiedad de ortonormalidad) para $j, k = 1, 2, \dots, p$ es:

$$\sum_{x=1}^p \mathbf{e}_{j(x)} \mathbf{e}_{k(x)} = \delta_{jk} \quad (A80)$$

Mientras que la versión matricial es:

$$\underline{\mathbf{E}}^T \underline{\mathbf{E}} = \underline{\mathbf{I}}_p \quad (A81)$$

Donde $\underline{\mathbf{I}}_p$ es la matriz identidad de dimensiones $(p \times p)$. De la ecuación (A81) podemos deducir que $\underline{\mathbf{E}}^T$ es la matriz inversa de $\underline{\mathbf{E}}$, de modo que también tenemos

$$\underline{\mathbf{E}} \underline{\mathbf{E}}^T = \underline{\mathbf{I}}_p \quad (A82)$$

Cuya versión escalar (propiedad de completitud) para $\{x, x' = 1, 2, \dots, p\}$ es:

$$\sum_{j=1}^p \mathbf{e}_{j(x)} \mathbf{e}_{j(x')} = \delta_{xx'} \quad (A83)$$

Una representación muy útil de $\underline{\mathbf{S}}$ se puede obtener de las ecuaciones (A77) y (A82).

$$\begin{aligned} \underline{\mathbf{S}} \underline{\mathbf{E}} &= \underline{\mathbf{E}} \underline{\mathbf{L}} \\ \underline{\mathbf{S}} \underline{\mathbf{E}} \underline{\mathbf{E}}^T &= \underline{\mathbf{E}} \underline{\mathbf{L}} \underline{\mathbf{E}}^T \end{aligned}$$

Aplicando la ecuación (A82)

$$\begin{aligned}\underline{\mathbf{S}} \underline{\mathbf{I}}_p &= \underline{\mathbf{E}} \underline{\mathbf{L}} \underline{\mathbf{E}}^T \\ \underline{\mathbf{S}} &= \underline{\mathbf{E}} \underline{\mathbf{L}} \underline{\mathbf{E}}^T\end{aligned}\quad (A84)$$

El cual se reduce a la forma vectorial

$$\underline{\mathbf{S}} = \sum_{j=1}^p l_j \underline{\mathbf{e}}_j \underline{\mathbf{e}}_j^T \quad (A85)$$

Cuya forma escalar para $\{x, x' = 1, 2, \dots, p\}$ es:

$$s_{(x, x')} = \sum_{j=1}^p l_j e_{j(x)} e_{j(x')} \quad (A86)$$

Representaciones básicas de los conjuntos de datos: Fórmulas de análisis y síntesis

La representación de la matriz de datos en términos de las componentes principales surge de la siguiente forma (usando A82):

$$\underline{\mathbf{Z}} = \underline{\mathbf{Z}} \underline{\mathbf{I}}_p = \underline{\mathbf{Z}} (\underline{\mathbf{E}} \underline{\mathbf{E}}^T) = (\underline{\mathbf{Z}} \underline{\mathbf{E}}) \underline{\mathbf{E}}^T \quad (A87)$$

Si definimos:

$$\underline{\mathbf{A}} = \underline{\mathbf{Z}} \underline{\mathbf{E}} \quad (n \times p) \quad (\text{análisis}) \quad (A88)$$

Entonces (A87) se convierte en:

$$\underline{\mathbf{Z}} = \underline{\mathbf{A}} \underline{\mathbf{E}}^T \quad (n \times p) \quad (\text{síntesis}) \quad (A89)$$

Las ecuaciones (A88) y (A89) son una generalización de las ecuaciones (A30) y (A29) vistas para el caso bidimensional.

$$\begin{aligned}\underline{\mathbf{A}} &= [\underline{\mathbf{a}}_1, \dots, \underline{\mathbf{a}}_p] \\ \underline{\mathbf{A}} &= \begin{bmatrix} a_{11} & \cdots & a_{p1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{pn} \end{bmatrix} \quad (\text{Matriz amplitud})\end{aligned} \quad (A90)$$

Los $\underline{\mathbf{a}}_j$ son los vectores amplitud, donde se ha definido sus componentes como:

$$\underline{\mathbf{a}}_j = [a_{j(1)}, \dots, a_{j(n)}]^T \quad ; \quad (j = 1, \dots, p)$$

Repaso y definición de las matrices

- Sea $\underline{\mathbf{E}}$ la matriz de autovectores de dimensión $(p \times p)$

$$\Rightarrow \underline{\mathbf{E}} = \begin{bmatrix} e_{1(1)} & \dots & \underbrace{e_{j(1)}}_{\underline{e}_j} & \dots & e_{p(1)} \\ \vdots & & \vdots & & \vdots \\ e_{1(p)} & \dots & \underbrace{e_{j(p)}}_{\underline{e}_j} & \dots & e_{p(p)} \end{bmatrix}_{(n \times p)}$$

Donde \underline{e}_j es un vector unitario con p elementos definido como:

$$\underline{e}_j = [e_{j(1)}, \dots, e_{j(p)}]^T, \quad j = 1, \dots, p$$

Por lo tanto $\underline{\mathbf{E}}$ puede ser definido como:

$$\underline{\mathbf{E}} = [\underline{e}_1, \dots, \underline{e}_j, \dots, \underline{e}_p]$$

- Sea $\underline{\mathbf{A}}$ la matriz amplitud de dimensión $(n \times p)$

$$\Rightarrow \underline{\mathbf{A}} = \begin{bmatrix} a_{1(1)} & \dots & \underbrace{a_{j(1)}}_{\underline{a}_j} & \dots & a_{p(1)} \\ \vdots & & \vdots & & \vdots \\ a_{1(n)} & \dots & \underbrace{a_{j(n)}}_{\underline{a}_j} & \dots & a_{p(n)} \end{bmatrix}_{(n \times p)}$$

Donde \underline{a}_j se define como:

$$\underline{a}_j = [a_{j(1)}, \dots, a_{j(n)}]^T, \quad j = 1, \dots, p$$

Por lo tanto $\underline{\mathbf{A}}$ puede ser definido como:

$$\underline{\mathbf{A}} = [\underline{a}_1, \dots, \underline{a}_j, \dots, \underline{a}_p]$$

- Sea $\underline{\mathbf{Z}}$ la matriz de datos centrados de dimensión $(n \times p)$

$$\underline{\mathbf{Z}} = \begin{bmatrix} z_{(1,1)} & \dots & z_{(1,x)} & \dots & z_{(1,p)} \\ \vdots & & \vdots & & \vdots \\ \boxed{z_{(t,1)} \quad \dots \quad z_{(t,x)} \quad \dots \quad z_{(t,p)}} & \dots & \dots & \dots & \dots \\ \vdots & & \vdots & & \vdots \\ z_{(n,1)} & \dots & z_{(n,x)} & \dots & z_{(n,p)} \end{bmatrix}_{(n \times p)} \} \underline{\mathbf{z}}_{(t)}^T$$

$$\underline{\mathbf{Z}} = \begin{bmatrix} z_{(1,1)} & \cdots & \boxed{z_{(1,x)}} & \cdots & z_{(1,p)} \\ \vdots & & \vdots & & \vdots \\ z_{(t,1)} & \cdots & \boxed{z_{(t,x)}} & \cdots & z_{(t,p)} \\ \vdots & & \vdots & & \vdots \\ z_{(n,1)} & \cdots & \boxed{z_{(n,x)}} & \cdots & z_{(n,p)} \end{bmatrix}_{n \times p}$$

$\underbrace{\hspace{1.5cm}}_{\mathbf{z}_{(*,x)}}$

Donde:

$$\boxed{\mathbf{z}_{(t)} = [z_{(t,1)}, \dots, z_{(t,x)}, \dots, z_{(t,p)}]^T}, \quad t = 1, \dots, n$$

$$\boxed{\mathbf{z}_{(*,x)} = [z_{(1,x)}, \dots, z_{(t,x)}, \dots, z_{(n,x)}]^T}, \quad x = 1, \dots, p$$

De la ecuación (A89) tenemos:

$$\Rightarrow \underline{\mathbf{Z}} = \underline{\mathbf{A}} \underline{\mathbf{E}}^T = \begin{bmatrix} a_{1(1)} & \cdots & a_{j(1)} & \cdots & a_{p(1)} \\ \vdots & & \vdots & & \vdots \\ a_{1(t)} & \cdots & a_{j(t)} & \cdots & a_{p(t)} \\ \vdots & & \vdots & & \vdots \\ a_{1(n)} & \cdots & a_{j(n)} & \cdots & a_{p(n)} \end{bmatrix}_{n \times p} \begin{bmatrix} e_{1(1)} & \cdots & e_{1(x)} & \cdots & e_{1(p)} \\ \vdots & & \vdots & & \vdots \\ e_{j(1)} & \cdots & e_{j(x)} & \cdots & e_{j(p)} \\ \vdots & & \vdots & & \vdots \\ e_{p(1)} & \cdots & e_{p(x)} & \cdots & e_{p(p)} \end{bmatrix}_{p \times p}$$

$$\Rightarrow \underline{\mathbf{Z}} = \begin{bmatrix} \sum_{j=1}^p a_{j(1)} e_{j(1)} & \cdots & \sum_{j=1}^p a_{j(1)} e_{j(x)} & \cdots & \sum_{j=1}^p a_{j(1)} e_{j(p)} \\ \vdots & & \vdots & & \vdots \\ \sum_{j=1}^p a_{j(t)} e_{j(1)} & \cdots & \sum_{j=1}^p a_{j(t)} e_{j(x)} & \cdots & \sum_{j=1}^p a_{j(t)} e_{j(p)} \\ \vdots & & \vdots & & \vdots \\ \sum_{j=1}^p a_{j(n)} e_{j(1)} & \cdots & \sum_{j=1}^p a_{j(n)} e_{j(x)} & \cdots & \sum_{j=1}^p a_{j(n)} e_{j(p)} \end{bmatrix}_{n \times p} \dots \dots \dots (\emptyset)$$

Si observamos la matriz (\emptyset) , esta puede ser escrita de la siguiente manera:

$$\underline{\mathbf{Z}} = \sum_{j=1}^p \underline{\mathbf{a}}_j \underline{\mathbf{e}}_j^T \quad (A91)$$

De las definiciones de $\mathbf{z}_{(t)}$ y $\mathbf{z}_{(*,x)}$ dadas anteriormente podemos señalar de la ecuación

(\emptyset) que:

$$\underline{\mathbf{z}}_{(t)} = \left[\sum_{j=1}^p a_{j(t)} e_{j(1)}, \dots, \sum_{j=1}^p a_{j(t)} e_{j(x)}, \dots, \sum_{j=1}^p a_{j(t)} e_{j(p)} \right]^T$$

$$\begin{aligned}\underline{\mathbf{z}}(\mathbf{t}) &= \sum_{j=1}^p a_{j(\mathbf{t})} \underbrace{[e_{j(1)}, \dots, e_{j(x)}, \dots, e_{j(p)}]}_{\underline{\mathbf{e}}_j}^T \\ \underline{\mathbf{z}}(\mathbf{t}) &= \underline{\mathbf{z}}(\mathbf{t}, *) = \sum_{j=1}^p a_{j(\mathbf{t})} \underline{\mathbf{e}}_j\end{aligned}\quad (A92a)$$

De la misma manera:

$$\begin{aligned}\underline{\mathbf{z}}(*, \mathbf{x}) &= \left[\sum_{j=1}^p a_{j(1)} e_{j(\mathbf{x})}, \dots, \sum_{j=1}^p a_{j(\mathbf{t})} e_{j(\mathbf{x})}, \dots, \sum_{j=1}^p a_{j(n)} e_{j(\mathbf{x})} \right]^T \\ \underline{\mathbf{z}}(*, \mathbf{x}) &= \sum_{j=1}^p \underbrace{[a_{j(1)}, \dots, a_{j(\mathbf{t})}, \dots, a_{j(n)}]}_{\underline{\mathbf{a}}_j}^T e_{j(\mathbf{x})} \\ \underline{\mathbf{z}}(*, \mathbf{x}) &= \underline{\mathbf{z}}(\mathbf{x}) = \sum_{j=1}^p \underline{\mathbf{a}}_j e_{j(\mathbf{x})}\end{aligned}\quad (A92b)$$

Para encontrar un elemento de la matriz $\underline{\mathbf{z}}$ (versión escalar de (A89)) sólo debemos observar la matriz (\emptyset)

$$\mathbf{z}(\mathbf{t}, \mathbf{x}) = \sum_{j=1}^p a_{j(\mathbf{t})} e_{j(\mathbf{x})} \quad (\text{síntesis}) \quad (A93)$$

Ahora intentaremos encontrar la versión vectorial y escalar de la ecuación (A88)

$$\begin{aligned}\underline{\mathbf{A}} &= \underline{\mathbf{Z}} \underline{\mathbf{E}} \\ [\underline{\mathbf{a}}_1, \dots, \underline{\mathbf{a}}_j, \dots, \underline{\mathbf{a}}_p] &= \underline{\mathbf{Z}} [\underline{\mathbf{e}}_1, \dots, \underline{\mathbf{e}}_j, \dots, \underline{\mathbf{e}}_p]\end{aligned}$$

Entonces la versión vectorial de (A88) es:

$$\underline{\mathbf{a}}_j = \underline{\mathbf{Z}} \underline{\mathbf{e}}_j \quad (A94)$$

La ecuación (A94) se puede expresar de manera matricial como:

$$\begin{bmatrix} a_{j(1)} \\ \vdots \\ a_{j(\mathbf{t})} \\ \vdots \\ a_{j(n)} \end{bmatrix} = \begin{bmatrix} z_{(1,1)} & \cdots & z_{(1,x)} & \cdots & z_{(1,p)} \\ \vdots & & \vdots & & \vdots \\ z_{(t,1)} & \cdots & z_{(t,x)} & \cdots & z_{(t,p)} \\ \vdots & & \vdots & & \vdots \\ z_{(n,1)} & \cdots & z_{(n,x)} & \cdots & z_{(n,p)} \end{bmatrix} \underbrace{\begin{bmatrix} e_{j(1)} \\ \vdots \\ e_{j(x)} \\ \vdots \\ e_{j(p)} \end{bmatrix}}_{\underline{\mathbf{e}}_j}$$

La versión escalar de (A88) es:

$$a_{j(\mathbf{t})} = \underbrace{[z_{(t,1)}, \dots, z_{(t,x)}, \dots, z_{(t,p)}]}_{\underline{\mathbf{z}}(\mathbf{t})^T} \underline{\mathbf{e}}_j$$

$$a_{j(t)} = \underline{z}_{(t)}^T \underline{e}_j \quad (A95a)$$

O si se prefiere con sumatorias se puede expresar como:

$$a_{j(t)} = z_{(t,1)}e_{j(1)} + \cdots + z_{(t,x)}e_{j(x)} + \cdots + z_{(t,p)}e_{j(p)}$$

$$a_{j(t)} = \sum_{x=1}^p z_{(t,x)}e_{j(x)} \quad (A95b)$$

La propiedad del análisis PCA

Derivemos una propiedad fundamental de las componentes principales de un conjunto de datos $\underline{\mathbf{Z}}$ de dimensión $(n \times p)$. Las series correspondientes a las componentes principales no están correlacionadas por pares. Esto se deduce de la definición de $\underline{\mathbf{A}}$ dada en (A88).

$$\underline{\mathbf{A}} = \underline{\mathbf{Z}} \underline{\mathbf{E}}$$

$$\Rightarrow \underline{\mathbf{A}}^T \underline{\mathbf{A}} = (\underline{\mathbf{Z}} \underline{\mathbf{E}})^T (\underline{\mathbf{Z}} \underline{\mathbf{E}})$$

$$\underline{\mathbf{A}}^T \underline{\mathbf{A}} = (\underline{\mathbf{Z}} \underline{\mathbf{E}})^T (\underline{\mathbf{Z}} \underline{\mathbf{E}})$$

$$\underline{\mathbf{A}}^T \underline{\mathbf{A}} = \underline{\mathbf{E}}^T \underbrace{\underline{\mathbf{Z}}^T \underline{\mathbf{Z}}}_{\underline{\mathbf{S}}} \underline{\mathbf{E}}$$

$$\underline{\mathbf{A}}^T \underline{\mathbf{A}} = \underline{\mathbf{E}}^T \underbrace{\underline{\mathbf{S}} \underline{\mathbf{E}}}_{(A.77)}$$

$$\underline{\mathbf{A}}^T \underline{\mathbf{A}} = \underline{\mathbf{E}}^T \underbrace{\underline{\mathbf{E}}^T \underline{\mathbf{E}}}_{\underline{\mathbf{I}}_p} \underline{\mathbf{L}}$$

$$\therefore \underline{\mathbf{A}}^T \underline{\mathbf{A}} = \underline{\mathbf{L}} \quad (A96)$$

La versión vectorial de (A96) se expresa como:

$$\underline{\mathbf{a}}_j^T \underline{\mathbf{a}}_k = l_j \delta_{jk} \quad ; \quad j, k = 1, 2, \dots, p \quad (A97)$$

Que a su vez tiene la siguiente forma escalar:

$$\sum_{t=1}^n a_{j(t)} a_{k(t)} = l_j \delta_{jk} \quad (A97a)$$

La ecuación (A96) es la generalización de (A33). Note que $l_j^{1/2}$ tiene las mismas dimensiones que los elementos de la matriz de datos $\underline{\mathbf{Z}}$. La ecuación (A96) es llamada la propiedad PCA del conjunto de datos $\underline{\mathbf{Z}}$.

Usando la ecuación (A96) podemos resolver la ecuación (A89) para $\underline{\mathbf{E}}$. Si multiplicamos por la izquierda $\underline{\mathbf{A}}^T$ en cada lado de (A89) obtenemos:

$$\underline{\mathbf{A}}^T \underline{\mathbf{Z}} = \underbrace{\underline{\mathbf{A}}^T \underline{\mathbf{A}}}_{\underline{\mathbf{L}}} \underline{\mathbf{E}}^T$$

$$\underline{\mathbf{A}}^T \underline{\mathbf{Z}} = \underline{\mathbf{L}} \underline{\mathbf{E}}^T$$

Calculando la transpuesta:

$$(\underline{\mathbf{A}}^T \underline{\mathbf{Z}})^T = (\underline{\mathbf{L}} \underline{\mathbf{E}}^T)^T$$

$$\underline{\mathbf{Z}}^T \underline{\mathbf{A}} = \underline{\mathbf{E}} \underbrace{\underline{\mathbf{L}}^T}_{\underline{\mathbf{L}}}$$

$$\underline{\mathbf{Z}}^T \underline{\mathbf{A}} = \underline{\mathbf{E}} \underline{\mathbf{L}}$$

Suponiendo que todos los valores de $\underline{\mathbf{L}}$ son positivos encontramos una representación de $\underline{\mathbf{E}}$ en función de $\underline{\mathbf{Z}}$ y $\underline{\mathbf{A}}$.

$$\underline{\mathbf{E}} = \underline{\mathbf{Z}}^T \underline{\mathbf{A}} \underline{\mathbf{L}}^{-1} \quad (\text{A98})$$

En forma vectorial:

$$\underline{\mathbf{e}}_j = l_j^{-1} \underline{\mathbf{Z}}^T \underline{\mathbf{a}}_j \quad (\text{A98a})$$

O en forma escalar (en donde $\{j = 1, 2, \dots, p\}$ y $\{x = 1, 2, \dots, p\}$):

$$e_{j(x)} = l_j^{-1} \sum_{t=1}^n z_{(t,x)} a_{j(t)} = l_j^{-1} \underline{\mathbf{z}}_{(x)}^T \underline{\mathbf{a}}_j \quad (\text{A98b})$$

Propiedades de Segundo Orden del análisis PCA: La dispersión total

Algunas propiedades útiles de 2do orden del análisis PCA de los datos involucran dispersión (es decir varianzas) del conjunto de datos y varias correlaciones. Por ejemplo si imaginamos a $\{\mathbf{z}_{(t,x)}: t = 1, 2, \dots, n\}$ como una serie de tiempo de " n " valores en un punto fijo " x " en el espacio, entonces la dispersión ($= (n - 1) \times \text{varianza}$) de esta serie temporal centrada en " t " usando (A93) y (A97a) es:

$$\begin{aligned} \sum_{t=1}^n \mathbf{z}_{(t,x)}^2 &= \sum_{t=1}^n \left[\sum_{j=1}^p a_{j(t)} e_{j(x)} \right]^2 \\ \Rightarrow \sum_{t=1}^n \mathbf{z}_{(t,x)}^2 &= \sum_{j=1}^p \underbrace{\left[\sum_{t=1}^n a_{j(t)}^2 \right]}_{(\text{A97a})} e_{j(x)}^2 \\ \Rightarrow \sum_{t=1}^n \mathbf{z}_{(t,x)}^2 &= \sum_{j=1}^p l_j e_{j(x)}^2 \end{aligned} \quad (\text{A99})$$

En donde se observa que (A99) es un caso especial de la propiedad de 2do orden. Analizando esta vez $x \neq x'$ donde $\{x, x' = 1, 2 \dots \dots, p\}$ obtenemos:

$$\sum_{t=1}^n \mathbf{z}_{(t,x)} \mathbf{z}_{(t,x')} = \sum_{j=1}^p l_j \mathbf{e}_{j(x)} \mathbf{e}_{j(x')} = \mathbf{s}_{(x,x')} \quad (A99a)$$

El cual es por supuesto igual a la ecuación (A86).

La dispersión total $(= (n-1) \times \text{varianza total})$ del conjunto de datos se encuentra usando las ecuaciones (A99) y (A80).

$$\begin{aligned} \sum_{x=1}^p \sum_{t=1}^n \mathbf{z}_{(t,x)}^2 &= \sum_{x=1}^p \sum_{j=1}^p l_j \mathbf{e}_{j(x)}^2 \\ \sum_{x=1}^p \sum_{t=1}^n \mathbf{z}_{(t,x)}^2 &= \sum_{j=1}^p l_j \underbrace{\sum_{x=1}^p \mathbf{e}_{j(x)}^2}_{(A80)} \\ \Rightarrow \sum_{x=1}^p \sum_{t=1}^n \mathbf{z}_{(t,x)}^2 &= \sum_{j=1}^p l_j \end{aligned} \quad (A100)$$

La ecuación (A100) muestra que la dispersión total de los datos es dado por la suma de los autovalores de $\underline{\mathbf{S}}$. Además de (A97a) se tiene que:

$$\sum_{j=1}^p l_j = \sum_{j=1}^p \sum_{t=1}^n a_{j(t)}^2 \quad (A100a)$$

La ecuación (A100a) indica que la dispersión total de los datos es invariante bajo la transformación del marco de coordenadas $\underline{\mathbf{E}}$ (ver A35).

La ecuación (A99) muestra como la varianza temporal de $\mathbf{z}_{(t,x)}$ en un punto espacial " \mathbf{x} " es ponderado por las componentes $\mathbf{e}_{j(x)}^2$ los cuales para un " \mathbf{x} " fijo mediante (A83) suma 1 sobre todos los valores de " \mathbf{j} ". Por otro lado para cada " \mathbf{j} " fijo, aquellos lugares " \mathbf{x} " en el espacio donde los $\mathbf{e}_{j(x)}^2$ son grandes tienen las mayores proporciones de varianza de las series temporales $\underline{\mathbf{z}}_{(*,x)}$.

Los autovalores l_j de $\underline{\mathbf{S}}$ están por tanto estrechamente conectados con la dispersión del conjunto de datos $\underline{\mathbf{z}}$. De hecho al retornar a la ecuación (A64) encontramos que para cada autovector $\underline{\mathbf{e}}_j$ usando (A74) y (A78) obtenemos:

$$\psi(\underline{e}_j) = \underline{e}_j^T \underbrace{\underline{S} \underline{e}_j}_{(A74)} \geq 0$$

$$\psi(\underline{e}_j) = \underline{e}_j^T \underline{l}_j \underline{e}_j \geq 0$$

$$\psi(\underline{e}_j) = \underline{l}_j \underbrace{\underline{e}_j^T \underline{e}_j}_{(A78)} \geq 0$$

$$\Rightarrow \underline{l}_j \geq 0 \quad (A101)$$

Por lo tanto el autovalor \underline{l}_j (no negativo) de \underline{S} es la dispersión de \underline{z} a lo largo de la dirección \underline{e}_j en " E_p " (ver A10 y A49).

La contraparte de (A99) se encuentra usando (A93) y (A80) para $t = 1, 2, \dots, n$.

$$\begin{aligned} \sum_{x=1}^p \mathbf{z}_{(t,x)}^2 &= \sum_{x=1}^p \left[\sum_{j=1}^p a_{j(t)} e_{j(x)} \right]^2 \\ \Rightarrow \sum_{x=1}^p \mathbf{z}_{(t,x)}^2 &= \sum_{j=1}^p a_{j(t)}^2 \underbrace{\sum_{x=1}^p e_{j(x)}^2}_{(A80)} \\ &\Rightarrow \sum_{x=1}^p \mathbf{z}_{(t,x)}^2 = \sum_{j=1}^p a_{j(t)}^2 \\ &\Rightarrow \sum_{x=1}^p \mathbf{z}_{(t,x)}^2 = \sum_{j=1}^p \underline{l}_j \alpha_{j(t)}^2 \end{aligned} \quad (A102)$$

Este es el caso especial (relación de Rayleigh) de la siguiente propiedad básica de 2do orden para $t, t' = 1, 2, \dots, n$:

$$\begin{aligned} \sum_{x=1}^p \mathbf{z}_{(t,x)} \mathbf{z}_{(t',x)} &= \sum_{j=1}^p a_{j(t)} a_{j(t')} \\ \Rightarrow \sum_{x=1}^p \mathbf{z}_{(t,x)} \mathbf{z}_{(t',x)} &= \sum_{j=1}^p \underline{l}_j \alpha_{j(t)} \alpha_{j(t')} \end{aligned} \quad (A102a)$$

Esta es la contraparte PCA de la relación de Parseval, la cual es útil en el análisis armónico y álgebra lineal. Aquí para cada $\underline{l}_j > 0$ se ha introducido la amplitud normalizada (componentes principales adimensionales).

$$\alpha_{j(t)} = a_{j(t)} / \underline{l}_j^{1/2} \quad (A103)$$

Descomposición de valores singulares (SVD) de un conjunto de datos

El método SVD es fácilmente obtenido de la fórmula de síntesis de PCA. Definamos:

$$\underline{\mathbf{A}}' = [\underline{\alpha}_1, \dots, \underline{\alpha}_p] \quad ; \quad (n \times p) \quad (A104)$$

donde:

$$\underline{\alpha}_j = [\alpha_{j(1)}, \dots, \alpha_{j(n)}]^T \quad ; \quad j = 1, \dots, p \quad (A105)$$

Por lo tanto tenemos que:

$$\underline{\mathbf{A}} = \underline{\mathbf{A}}' \underline{\mathbf{L}}^{1/2} \quad , \quad \underline{\mathbf{L}}^{1/2} = \text{diag}[l_1^{1/2}, \dots, l_p^{1/2}] \quad (A106)$$

Y (A96) se convierte en:

$$\begin{aligned} \underline{\mathbf{A}}^T \underline{\mathbf{A}} &= \underline{\mathbf{L}} \\ \underline{\mathbf{L}}^{1/2} \underline{\mathbf{A}}'^T \underline{\mathbf{A}}' \underline{\mathbf{L}}^{1/2} &= \underline{\mathbf{L}} \\ \therefore \underline{\mathbf{A}}'^T \underline{\mathbf{A}}' &= \underline{\mathbf{I}}_p \end{aligned} \quad (A107)$$

Ahora encontremos la contraparte matricial, vectorial y escalar de (A89), (A92a) y (A93).

De (A89):

$$\begin{aligned} \underline{\mathbf{Z}} &= \underline{\mathbf{A}} \underline{\mathbf{E}}^T \\ \underline{\mathbf{Z}} &= \underline{\mathbf{A}}' \underline{\mathbf{L}}^{1/2} \underline{\mathbf{E}}^T \end{aligned} \quad (n \times p) \quad (A108)$$

De (A92a):

$$\underline{\mathbf{z}}_{(t)} = \sum_{j=1}^p l_j^{1/2} \alpha_{j(t)} \underline{\mathbf{e}}_j \quad (A109)$$

De (A93) para $t = 1, 2, \dots, n$ y $x = 1, 2, \dots, p$:

$$\underline{\mathbf{z}}_{(t,x)} = \sum_{j=1}^p l_j^{1/2} \alpha_{j(t)} \underline{\mathbf{e}}_{j(x)} \quad (A110)$$

La ecuación (A108) constituye una forma de lo que se denomina SVD de un conjunto de datos $\underline{\mathbf{Z}}$. Note que $\underline{\mathbf{A}}'$ en (A107) tiene su contraparte para $\underline{\mathbf{E}}$ en (A81).

Propiedades de 2do orden del análisis PCA: Correlaciones

Retornemos a la discusión sobre las propiedades de 2do orden del análisis PCA del conjunto de datos y consideremos la correlación de las series temporales $\underline{\mathbf{z}}_{(t,x)}$ con las

series $\mathbf{a}_{j(t)}$ donde $t = 1, \dots, n$. Dado que cada uno de estos conjuntos está centrado en el tiempo, su covarianza mediante (A93) y (A97a) es:

$$\begin{aligned}
 \sum_{t=1}^n \underbrace{\mathbf{z}_{(t,x)}}_{(A93)} \mathbf{a}_{j(t)} &= \sum_{t=1}^n \left[\sum_{k=1}^p \mathbf{a}_{k(t)} \mathbf{e}_{k(x)} \right] \mathbf{a}_{j(t)} \\
 &= \sum_{k=1}^p \underbrace{\left[\sum_{t=1}^n \mathbf{a}_{k(t)} \mathbf{a}_{j(t)} \right]}_{(A97a)} \mathbf{e}_{k(x)} \\
 &= \sum_{k=1}^p l_j \delta_{jk} \mathbf{e}_{k(x)} \\
 \sum_{t=1}^n \mathbf{z}_{(t,x)} \mathbf{a}_{j(t)} &= l_j \mathbf{e}_{j(x)} \quad ; x = 1, 2, \dots, p \quad (A111)
 \end{aligned}$$

El cuadrado de su correlación usando (A97) y (A99) es entonces:

$$\begin{aligned}
 \mathbf{r}_{\mathbf{za}(j,x)}^2 &= \frac{\left[\sum_{t=1}^n \mathbf{z}_{(t,x)} \mathbf{a}_{j(t)} \right]^2}{\underbrace{\left[\sum_{t=1}^n \mathbf{z}_{(t,x)}^2 \right]}_{(A99)} \underbrace{\left[\sum_{t=1}^n \mathbf{a}_{j(t)}^2 \right]}_{(A97)}} \\
 \mathbf{r}_{\mathbf{za}(j,x)}^2 &= \frac{l_j^2 \mathbf{e}_{j(x)}^2}{\left[\sum_{j=1}^p l_j \mathbf{e}_{j(x)}^2 \right] [l_j]} \\
 \mathbf{r}_{\mathbf{za}(j,x)}^2 &= \frac{l_j \mathbf{e}_{j(x)}^2}{\sum_{j=1}^p l_j \mathbf{e}_{j(x)}^2} \geq 0 \quad (A112)
 \end{aligned}$$

para $j = 1, 2, \dots, p$ y $x = 1, 2, \dots, p$. El estadístico $\mathbf{r}_{\mathbf{za}(j,x)}^2$ da una medida de la correlación entre las series temporal de datos originales $\mathbf{z}_{(*,x)}$ en un punto "x" y la j-ésima componente principal \mathbf{a}_j . El estadístico tiene la propiedad aditiva en cada punto espacial $x = 1, 2, \dots, p$:

$$\sum_{j=1}^p \mathbf{r}_{\mathbf{za}(j,x)}^2 = 1 \quad (A113)$$

Uno puede imaginar un mapa de $\mathbf{r}_{\mathbf{za}(j,x)}^2$ graficado como una función de la posición "x" para cada índice "j". En aquellos lugares "x" donde la superficie del mapa es alta (valores grandes) existe relativamente gran correlación entre las variables \mathbf{a}_j y $\mathbf{z}_{(*,x)}$. Notar que

$\mathbf{r}_{\mathbf{za}(j,x)}^2$ tiene otra interpretación usando (A99): “Esta da una cantidad fraccional de la varianza de $\mathbf{z}_{(*,x)}$ en cada "x" asociada con el j-ésimo modo \mathbf{e}_j ”

En contraparte con la ecuación (A11), podemos preguntar acerca de la correlación entre el conjunto de datos, es decir las series espaciales $\{z_{(t,x)}: x = 1, \dots, p\}$ y el conjunto $\{e_{j(x)}: x = 1, \dots, p\}$. Aunque estos conjuntos de datos no estén centrados en sus promedios sobre "x", podemos calcular su producto interno sobre el rango "x".

$$\sum_{x=1}^p z_{(t,x)} e_{j(x)} = a_{j(t)} \quad ; t = 1, \dots, n \quad (A114)$$

El cual es por supuesto la ecuación (A95b). En analogía a (A112) y usando las ecuaciones (A102) y (A103) se obtiene:

$$\begin{aligned} \sum_{x=1}^p \underbrace{z_{(t,x)}}_{(A93)} e_{j(x)} &= \sum_{x=1}^p \left[\sum_{k=1}^p a_{k(t)} e_{k(x)} \right] e_{j(x)} \\ &= \sum_{k=1}^p \left[\underbrace{\sum_{x=1}^p e_{k(x)} e_{j(x)}}_{(A80)} \right] a_{k(t)} \\ &= \sum_{k=1}^p \delta_{jk} a_{k(t)} \\ &= a_{j(t)} \\ \sum_{x=1}^p z_{(t,x)} e_{j(x)} &= l_j^{1/2} \alpha_{j(t)} \\ \Rightarrow \mathbf{r}_{\mathbf{ze}(j,t)}^2 &= \frac{[\sum_{x=1}^p z_{(t,x)} e_{j(x)}]^2}{\left[\sum_{x=1}^p \underbrace{z_{(t,x)}^2}_{(A93)} \right] \underbrace{[\sum_{x=1}^p e_{j(x)}^2]}_1} \\ \Rightarrow \mathbf{r}_{\mathbf{ze}(j,t)}^2 &= \frac{l_j \alpha_{j(t)}^2}{\sum_{x=1}^p [\sum_{j=1}^p a_{j(t)}^2 e_{j(x)}^2]} \\ \Rightarrow \mathbf{r}_{\mathbf{ze}(j,t)}^2 &= \frac{l_j \alpha_{j(t)}^2}{\sum_{x=1}^p [\sum_{j=1}^p a_{j(t)}^2 e_{j(x)}^2]} = \frac{l_j \alpha_{j(t)}^2}{\sum_{j=1}^p \underbrace{[\sum_{x=1}^p e_{j(x)}^2]}_{=1} a_{j(t)}^2} = \frac{l_j \alpha_{j(t)}^2}{\sum_{j=1}^p a_{j(t)}^2} \end{aligned}$$

$$\Rightarrow \mathbf{r}_{\mathbf{ze}(j,t)}^2 = \frac{l_j \alpha_{j(t)}^2}{\sum_{j=1}^p l_j \alpha_{j(t)}^2} \geq 0 \quad (\text{A115})$$

para $j = 1, 2, \dots, p$ y $t = 1, 2, \dots, n$. El estadístico $\mathbf{r}_{\mathbf{ze}(j,t)}^2$ da una medida de la correlación entre el mapa de datos originales $\mathbf{z}_{(t,*)}$ en un tiempo "t" y el j-ésimo vector principal \mathbf{e}_j considerado también como un mapa espacial. El estadístico tiene la propiedad aditiva.

$$\sum_{j=1}^p \mathbf{r}_{\mathbf{ze}(j,t)}^2 = 1 \quad (\text{A116})$$

Para cada tiempo $t = 1, 2, \dots, n$. Uno puede imaginar una curva de $\mathbf{r}_{\mathbf{ze}(j,t)}^2$ graficada como una función del tiempo para cada índice "j". En aquellos tiempos "t" donde la curva es alta, existe relativamente gran correlación sobre el espacio entre el mapa espacial de \mathbf{e}_j y $\mathbf{z}_{(t,*)}$. Uno puede también considerar graficar las correlaciones $\mathbf{r}_{\mathbf{ze}}$ y $\mathbf{r}_{\mathbf{za}}$, en lugar de sus cuadrados, de modo que el signo de las correlaciones pueda ser visto; sin embargo las propiedades aditivas como (A113) y (A115) no se cumplen.

Propiedad del análisis PCA

Retornamos para examinar con más detalle la propiedad (A96) del conjunto de datos \mathbf{Z} de dimensiones $n \times p$. Una manera de colocar esta importante propiedad de \mathbf{Z} en perspectiva es dar a \mathbf{Z} otra representación por medio de un conjunto ortonormal arbitrario de vectores base $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_j, \dots, \mathbf{u}_p]$ de E_p en el cual los "n" vectores $\mathbf{z}_{(t)}$, $t = 1, 2, \dots, n$ existen. Así por definición de \mathbf{U} :

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}_p \quad (\text{A117})$$

O también:

$$\mathbf{U} \mathbf{U}^T = \mathbf{I}_p \quad (\text{A118})$$

Por lo tanto podemos escribir como en (A87)

$$\mathbf{Z} = \mathbf{Z} \mathbf{I}_p = \mathbf{Z} (\mathbf{U} \mathbf{U}^T) = (\mathbf{Z} \mathbf{U}) \mathbf{U}^T \quad (\text{A119})$$

y definimos una matriz de amplitud

$$\mathbf{B} = \mathbf{Z} \mathbf{U} \quad (n \times p) \quad (\text{A120})$$

Donde

$$\underline{\mathbf{B}} = [\underline{\mathbf{b}}_1, \dots, \underline{\mathbf{b}}_p] \quad (\text{A121})$$

En el que

$$\underline{\mathbf{b}}_j = [b_{j(1)}, \dots, b_{j(n)}]^T \quad ; \quad (j = 1, \dots, p)$$

Por lo tanto $\underline{\mathbf{Z}}$ tiene sus representaciones asociadas en los tres niveles

$$\left. \begin{aligned} \underline{\mathbf{Z}} &= \underline{\mathbf{B}} \underline{\mathbf{U}}^T && (\text{matriz}) \\ \underline{\mathbf{z}}_{(t)} &= \sum_{j=1}^p b_{j(t)} \underline{\mathbf{u}}_j && (\text{vector}) \\ z_{(t,x)} &= \sum_{j=1}^p b_{j(t)} u_{j(x)} && (\text{escalar}) \end{aligned} \right\} \quad (\text{A122})$$

De esta manera $\underline{\mathbf{z}}_{(t)}$, $t = 1, 2, \dots, n$ puede ser representado por una combinación lineal de vectores ortonormales $\underline{\mathbf{u}}_j$ en E_p . De hecho hay un número infinito de tales representaciones de $\underline{\mathbf{z}}_{(t)}$, uno por cada $\underline{\mathbf{U}}$. Sin embargo si $\underline{\mathbf{U}} \neq \underline{\mathbf{E}}$, donde $\underline{\mathbf{E}}$ es una matriz de autovectores (A75) definido por la matriz de dispersión $\underline{\mathbf{S}}$ de los elementos de $\underline{\mathbf{Z}}$, y los autovalores asociados de $\underline{\mathbf{S}}$ son distintos $l_1 > l_2 > l_3 > \dots > l_p > 0$, entonces la propiedad (A96) no es válida para $\underline{\mathbf{B}}$. En este sentido la propiedad (A96) caracteriza algebraicamente el PCA (la condición del l_j es mantener el \mathbf{e}_j asociado definido de forma única).

Propiedad asintótica del análisis PCA y sistemas dinámicos

La propiedad (A96) de $\underline{\mathbf{Z}}$ resulta tener profundas conexiones con varios modelos de procesos físicos estudiados en meteorología y oceanografía. Varios conjuntos de datos generados mediante soluciones de ecuaciones lineales o ecuaciones diferenciales parciales exhiben la propiedad (A96) en el límite de grandes tamaños de muestra " n ". Cuando éste es el caso los autovectores del conjunto de datos se asemejan a los modos espaciales ortogonales teóricos de las soluciones. De este modo las "funciones ortogonales empíricas" surgen con un significado físico definido.

Supongamos que los datos matriciales $\underline{\mathbf{Z}}$ de dimensión $(n \times p)$ centradas en el tiempo han sido generadas por un modelo lineal determinístico. Podemos representar los puntos $\underline{\mathbf{z}}_{(t)}$ en

E_p en cualquiera de los niveles mostrados en (A122). Los vectores $\underline{\mathbf{u}}_j$ en $\underline{\mathbf{U}}$ son los auto-movimientos del sistema dinámico. Si construimos $\underline{\mathbf{B}}$ a partir de $\underline{\mathbf{Z}}$ y $\underline{\mathbf{U}}$ como en (A120), no podemos esperar generalmente que $\underline{\mathbf{B}}^T \underline{\mathbf{B}}$ sea una matriz diagonal para muestras finitas de tamaño " n ". Sin embargo puede ocurrir que $(n-1)^{-1} \underline{\mathbf{B}}^T \underline{\mathbf{B}}$ en el límite de un n grande sea una matriz diagonal $\underline{\mathbf{M}}$ de dimensiones $(p \times p)$. Por la ecuación matricial en (A122) esto implica que existe una matriz de covarianza $\underline{\mathbf{C}}$ de dimensiones $(p \times p)$ tal que:

$$\begin{aligned}\underline{\mathbf{C}} &= \lim_{n \rightarrow \infty} (n-1)^{-1} \underline{\mathbf{Z}}^T \underline{\mathbf{Z}} \\ \underline{\mathbf{C}} &= \underline{\mathbf{U}} \left\{ \lim_{n \rightarrow \infty} (n-1)^{-1} \underline{\mathbf{B}}^T \underline{\mathbf{B}} \right\} \underline{\mathbf{U}}^T \\ \underline{\mathbf{C}} &= \underline{\mathbf{U}} \underline{\mathbf{M}} \underline{\mathbf{U}}^T\end{aligned}\tag{A123}$$

De modo que:

$$\underline{\mathbf{C}} \underline{\mathbf{U}} = \underline{\mathbf{U}} \underline{\mathbf{M}}\tag{A124}$$

De esto podemos ver que $\underline{\mathbf{M}}$ es la matriz diagonal de autovalores y $\underline{\mathbf{U}}$ la matriz de autovectores de la matriz de covarianza $\underline{\mathbf{C}}$ en el límite. De esta manera los auto-movimientos $\underline{\mathbf{U}}$ del sistema dinámico se manifiestan como los autovectores del análisis PCA de los campos de salida de los sistemas dinámicos $\underline{\mathbf{Z}}$.

Siempre que:

$$\lim_{n \rightarrow \infty} (n-1)^{-1} \underline{\mathbf{B}}^T \underline{\mathbf{B}} = \lim_{n \rightarrow \infty} (n-1)^{-1} \underline{\mathbf{U}}^T \underline{\mathbf{Z}}^T \underline{\mathbf{Z}} \underline{\mathbf{U}} = \underline{\mathbf{M}}\tag{A125}$$

para un conjunto de datos $\underline{\mathbf{Z}}$ generados por el modelo, diremos que $\underline{\mathbf{Z}}$ tiene la propiedad asintótica PCA (para un $\underline{\mathbf{U}}$ fijo). En la práctica podríamos inferir la posibilidad de (A125) incrementando el tamaño de las muestras en $\underline{\mathbf{Z}}$. Por el momento se escribirá $\underline{\mathbf{Z}}_n$ para señalar que es una matriz de dimensiones $(n \times p)$ con " n " variable y " p " fijo. Sean $\underline{\mathbf{A}}_n$ y $\underline{\mathbf{E}}_n$ las matrices de amplitud y de autovectores asociadas a $\underline{\mathbf{Z}}_n$, de modo que de acuerdo a (A89) puede escribirse como $\underline{\mathbf{Z}}_n = \underline{\mathbf{A}}_n \underline{\mathbf{E}}_n^T$ para cualquier tamaño de muestra " n ". Bajo la hipótesis (A125) y la ecuación (A123) $\underline{\mathbf{C}}_n = (n-1)^{-1} \underline{\mathbf{Z}}_n^T \underline{\mathbf{Z}}_n$ se aproxima a su límite $\underline{\mathbf{C}}$. Por otra parte por los teoremas del álgebra lineal que dicen que las matrices de autovalores y autovectores de $\underline{\mathbf{C}}$ (es decir $\underline{\mathbf{L}}$ y $\underline{\mathbf{E}}$) varían continuamente con $\underline{\mathbf{C}}$, resulta que si (A125) es válido, entonces:

$$\lim_{n \rightarrow \infty} \underline{\mathbf{E}}_n = \underline{\mathbf{U}} \quad \text{y} \quad \lim_{n \rightarrow \infty} (n-1)^{-1} \underline{\mathbf{L}}_n = \underline{\mathbf{M}}\tag{A126}$$

Por lo tanto debemos ser capaces de detectar empíricamente la presencia de la propiedad asintótica del método para el conjunto de datos $\underline{\mathbf{Z}}$ producidos por el modelo, ya sea buscando la convergencia de $\underline{\mathbf{E}}_n$ y $(n-1)^{-1}\underline{\mathbf{L}}_n$ a sus límites, o simplemente la convergencia de $\underline{\mathbf{C}}_n$ cuando "n" aumenta.

PCA de compuestos espaciales de conjuntos de datos

Consideremos la aplicación del análisis PCA a campos de datos compuestos, empezando con el caso de conjuntos de datos compuestos espacialmente.

Supongamos por ejemplo que tenemos datos centrados en el tiempo de temperatura superficial del mar $z_{1(t,x)}$ para $t = 1, \dots, n$; $x = 1, \dots, p$ y datos centrados de presión al nivel del mar $z_{2(t,x)}$ sobre el mismo rango temporal "t", pero posiblemente en algún rango "x" diferente donde $x = p+1, \dots, p+q$. Por ejemplo posiciones $x = 1, \dots, p$ pueden extenderse sobre el Pacífico y las posiciones $x = p+1, \dots, p+q$ pueden extenderse sobre Norteamérica. Entonces el conjunto de datos asociados $\underline{\mathbf{Z}}$ será una matriz de dimensiones $n \times (p+q)$ y tendrá la forma compuesta $[\underline{\mathbf{Z}}_1 : \underline{\mathbf{Z}}_2]$, donde $\underline{\mathbf{Z}}_1$ es la matriz $(n \times p)$ de valores $z_{1(t,x)}$, y $\underline{\mathbf{Z}}_2$ es la matriz $(n \times q)$ de valores $z_{2(t,x)}$.

Al igual que en (A61) y (A62), el PCA de $\underline{\mathbf{Z}}$ comienza con la matriz de dispersión.

$$\underline{\mathbf{S}} = \underline{\mathbf{Z}}^T \underline{\mathbf{Z}} = \begin{bmatrix} \underline{\mathbf{Z}}_1^T \\ \vdots \\ \underline{\mathbf{Z}}_2^T \end{bmatrix} [\underline{\mathbf{Z}}_1 : \underline{\mathbf{Z}}_2]$$

$$\underline{\mathbf{S}} = \begin{bmatrix} \underline{\mathbf{Z}}_1^T \underline{\mathbf{Z}}_1 & \underline{\mathbf{Z}}_1^T \underline{\mathbf{Z}}_2 \\ \underline{\mathbf{Z}}_2^T \underline{\mathbf{Z}}_1 & \underline{\mathbf{Z}}_2^T \underline{\mathbf{Z}}_2 \end{bmatrix} \quad (A127)$$

Este $\underline{\mathbf{S}}$ surge a partir de una operación extremal sobre la dispersión $\psi(\underline{\mathbf{e}})$ a los largo de $\underline{\mathbf{e}}$ en el espacio euclidiano $(p+q)$ dimensional " E_{p+q} " de "n" puntos de datos de la forma $\underline{\mathbf{z}}(t) = [\underline{\mathbf{Z}}_1^T, \underline{\mathbf{Z}}_2^T]^T$, $t = 1, \dots, n$.

Tal como se ha visto en las anteriores discusiones, el análisis PCA es un análisis de varianza orientada, es decir que busca las varianzas principales del conjunto de datos y re-ensambla el conjunto de datos como en (A108)-(A110) alrededor de estas varianzas (en la forma de autovalores y autovectores de $\underline{\mathbf{S}}$).

Ahora si los valores de campo que comprenden \underline{Z}_2 tienen una mayor variabilidad que aquellos valores de \underline{Z}_1 , debido únicamente a un percance en la elección de las unidades para \underline{Z}_1 y \underline{Z}_2 , entonces un sesgo puede ser introducido en los diagnósticos o predicciones de campo basados en (A127). Para evitar accidentes causados por las unidades sobre la variabilidad del campo cuando dos o más campos de diferentes dimensiones aparecen juntos en el conjunto de datos compuestos \underline{Z} , se debería estandarizar los datos antes de formar \underline{Z} .

Así en vez de usar $z_{j(t,x)} = z'_{j(t,x)} - \bar{z}_{j(x)}$ uno debería usar $z_{j(t,x)} = \frac{z'_{j(t,x)} - \bar{z}_{j(x)}}{s_{j(x)}}$, para $j = 1, 2$; $x = 1, 2, \dots, p$ y $t = 1, 2, \dots, n$. Aquí $\bar{z}_{j(x)}$ es el promedio temporal de los datos no centrados $z'_{j(t,x)}$, y $s_{j(x)}$ es la desviación estándar asociada. Observamos que la representación PCA (A93) puede fácilmente ser aplicada a un $\mathbf{z}_{(t,x)}$ estandarizado (dejando de lado momentáneamente el subíndice "j" en $\mathbf{z}_{(t,x)}$).

$$\mathbf{z}_{(t,x)} = \sum_{j=1}^p a_{j(t)} \mathbf{e}_{j(x)}$$

Donde: $z_{(t,x)} = \frac{[z'_{(t,x)} - \bar{z}_{(x)}]}{s_{(x)}}$

Por lo tanto la representación de los datos primarios $z'_{(t,x)}$ es: $z'_{(t,x)} = \bar{z}_{(x)} + s_{(x)} z_{(t,x)}$

Donde:

$$\bar{z}_{(x)} = n^{-1} \sum_{t=1}^n z'_{(t,x)}$$

$$s_{(x)}^2 = (n-1)^{-1} \sum_{t=1}^n [z'_{(t,x)} - \bar{z}_{(x)}]^2$$

La dificultad puede surgir en la interpretación de las PCA de una matriz compuesta, tal como $\underline{Z} = [\underline{Z}_1 : \underline{Z}_2]$ definido anteriormente. Recordar cómo la ecuación (A73) fue derivada, requiriendo una identidad dimensionalmente correcta en la forma de (A71). Implícito en la derivación de (A73) está la suposición de que la matriz de dispersión tiene entradas de una sola dimensión. Por lo tanto si \underline{Z} es una matriz compuesta de la forma $[\underline{Z}_1 : \dots : \underline{Z}_m]$, donde \underline{Z}_j es una matriz $(n \times p_j)$ para $(j = 1, \dots, m)$, entonces todos estos \underline{Z}_j deben tener las mismas entradas dimensionadas; si no lo es, entonces cada \underline{Z}_j deberá ser estandarizada

(centrada y adimensionalizada) de alguna manera. El método de estandarización sugerido anteriormente no es sino uno de varios métodos diferentes que un investigador puede inventar; sin embargo desde un aspecto estadístico éste es el más útil y ampliamente aceptado.

Las fórmulas de síntesis para campos compuestos $\underline{Z} = [\underline{Z}_1 : \underline{Z}_2]$ pueden seguir los patrones de (A88) y (A89). Sea \underline{E} la matriz de autovectores (de dimensión $(p + q) \times (p + q)$) de \underline{S} en (A127). Desde el modo de construcción de $\underline{Z} = [\underline{Z}_1 : \underline{Z}_2]$ particionamos \underline{E} en la forma:

$$\underline{E} = [\underline{e}_1, \dots, \underline{e}_j, \dots, \underline{e}_{p+q}] = \begin{bmatrix} \underline{E}_1 \\ \dots \\ \underline{E}_2 \end{bmatrix}$$

Donde \underline{E}_1 es una matriz de dimensiones $p \times (p + q)$ y \underline{E}_2 de dimensiones $q \times (p + q)$, donde \underline{e}_j tiene la forma $\underline{e}_j = [e_{j(1)}, \dots, e_{j(p)}, e_{j(p+1)}, \dots, e_{j(p+q)}]^T$ para $j = 1, \dots, p + q$.

Por la propiedad de ortonormalidad de \underline{E} , en la que se tiene que $\underline{E} \underline{E}^T = \underline{I}_{p+q}$ podemos dividir \underline{Z} en dos factores (ver A87).

$$\underline{Z} = (\underline{Z} \underline{E}) \underline{E}^T$$

$$[\underline{Z}_1 : \underline{Z}_2] = \left\{ [\underline{Z}_1 : \underline{Z}_2] \begin{bmatrix} \underline{E}_1 \\ \dots \\ \underline{E}_2 \end{bmatrix} \right\} \begin{bmatrix} \underline{E}_1 \\ \dots \\ \underline{E}_2 \end{bmatrix}$$

Donde:

$$\begin{aligned} \underline{Z}_1 &= \underline{A} \underline{E}_1^T & (n \times p) \\ \underline{Z}_2 &= \underline{A} \underline{E}_2^T & (n \times q) \\ \underline{A} &= \underline{Z}_1 \underline{E}_1 + \underline{Z}_2 \underline{E}_2 & (n \times (p + q)) \end{aligned}$$

En la forma escalar estas fórmulas para $t = 1, \dots, n$ son:

$$\begin{aligned} z_{1(t,x)} &= \sum_{j=1}^{p+q} a_{j(t)} e_{j(x)} \quad ; \quad x = 1, \dots, p \\ z_{2(t,x)} &= \sum_{j=1}^{p+q} a_{j(t)} e_{j(x)} \quad ; \quad x = p + 1, \dots, p + q \\ a_{j(t)} &= \sum_{x=1}^p z_{1(t,x)} e_{j(x)} + \sum_{x=p+1}^{p+q} z_{2(t,x)} e_{j(x)} \quad ; \quad j = 1, \dots, p + q \end{aligned}$$

La propiedad (A96) vale para la presente matriz de amplitud que tiene dimensiones $n \times (p + q)$. Por lo tanto:

$$\begin{aligned}\underline{A}^T \underline{A} &= \underline{L} \\ \underline{L} &= \text{diag}[l_1, \dots \dots l_p] \\ l_1 &\geq l_2 \geq l_3 \dots \dots \geq l_{p+q}\end{aligned}$$

Por otra parte las demás propiedades de 2do orden discutidas anteriormente también se cumplen cuando se formulan adecuadamente.

El análisis precedente de los campos compuestos se generalizan al caso donde $\underline{Z} = [\underline{Z}_1 : \dots : \underline{Z}_m]$. Los \underline{Z}_j 's están definidos sobre el mismo dominio temporal, pero pueden estar sobre distintos dominios espaciales.

ANEXO 2: Repaso de algunos conceptos fundamentales

1. Temperatura

Cuando dos objetos poseen la misma temperatura sus propiedades no experimentarán cambios cuando son puestos en contacto. Por lo tanto la “Temperatura” será aquella propiedad que determine si dos a más sistemas se encuentran en equilibrio térmico. Cuando los sistemas en contacto se encuentren a diferentes temperaturas tendrá lugar un flujo de energía o flujo de calor. Muchos de los fenómenos climáticos tienen relación directa con la temperatura (por ejemplo la evaporación y la condensación). A continuación se describen los principales factores que regulan el comportamiento de la temperatura en la superficie terrestre.

a) **Latitud:** La radiación solar incidente varía a lo largo de la latitud (figura B1).

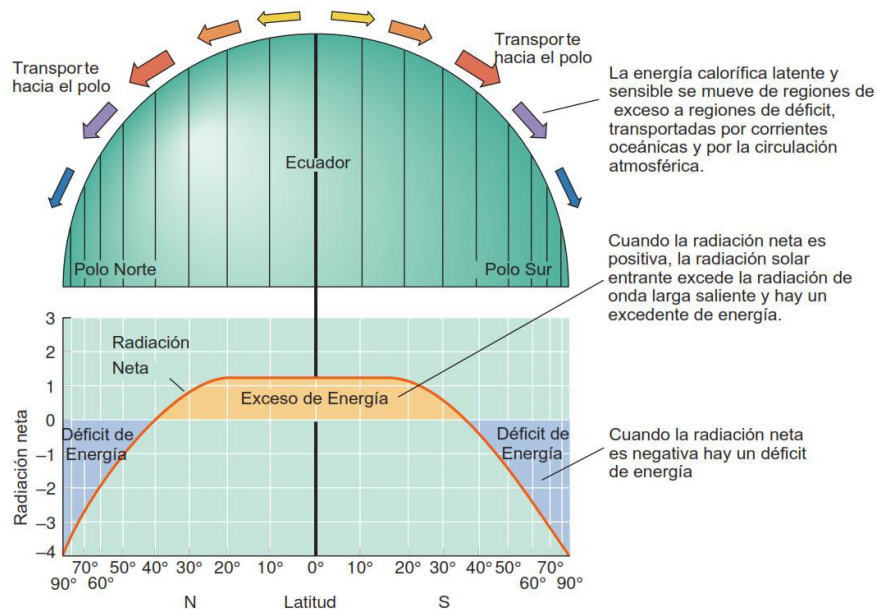


Figura B1. Radiación anual neta en la superficie de polo a polo (adaptado de Strahler, 2010).

- b) **Transparencia atmosférica:** Es la capacidad que tiene la atmósfera de la Tierra de transmitir la luz procedente de los objetos celestes. Depende principalmente de la cantidad de material absorbente presente (aerosoles, polvo, vapor de agua, gases contaminantes, etc).
- c) **Naturaleza de la Superficie:** Se tiene que distinguir el comportamiento de la superficie marítima y la superficie terrestre frente a la radiación. Por ejemplo la superficie marítima absorbe una gran cantidad de radiación (debe absorber una gran cantidad de calor para aumentar su temperatura). El calor se redistribuye dentro del

volumen de agua, empleándose una parte en los procesos de evaporación. Durante la noche, el mar emite radiación en grandes cantidades a la atmósfera, reteniéndose parte de la radiación debido al abundante vapor de agua existente en las capas de aire próximas a la superficie. Por todo esto la superficie marina es un regulador térmico.

- d) **Circulaciones atmosféricas y oceánicas:** Permiten la transmisión de calor mediante la advección entre distintas ubicaciones del planeta. En sus desplazamientos tanto las masas de aire como de agua transportan calor transmitiéndolo a los lugares que recorren.

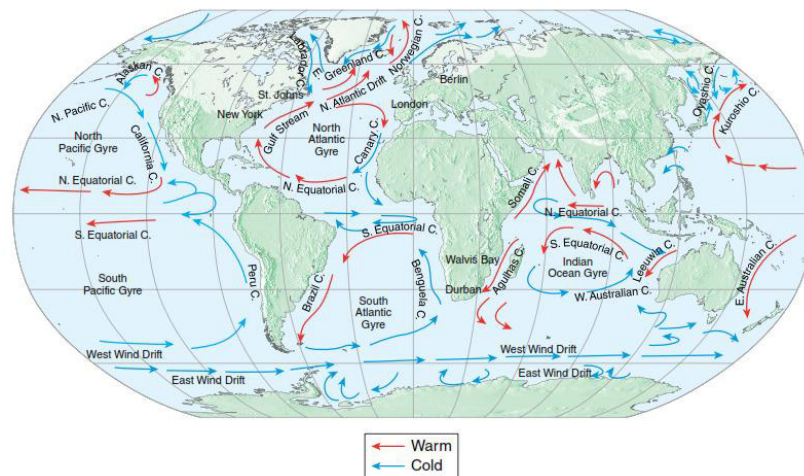


Figura B2. Circulación superficial del océano (Rohli and Vega, 2018).

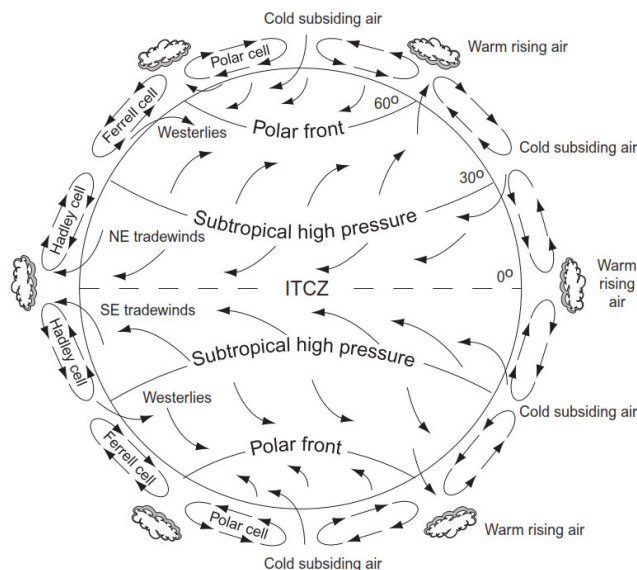


Figura B3. Circulación atmosférica (Chapin et al, 2011).

- e) **Altura sobre el nivel del mar:** Cuanto mayor sea la altura, menores las temperaturas registradas.
- f) **Topografía:** Las variaciones en la topografía también intervienen en la radiación solar que incide sobre un lugar, estableciendo, por ejemplo, diferencias entre solanas (lugar en donde inciden mucho los rayos solares.) y umbrías (sombra) de las cadenas montañosas.

2. Balance térmico de las aguas superficiales

Para un cuerpo de agua dado podemos establecer un balance calórico siguiendo la forma general de la ley de conservación. En otras palabras el calor ganado o perdido por unidad de tiempo por volumen de agua considerado es igual al flujo neto de calor a través de las paredes del volumen más la producción neta de calor en el interior del volumen por procesos de transformación de energía. El flujo de calor a través de las paredes del volumen puede ocurrir de dos maneras principales. Una de ellas es la difusión molecular y turbulenta, y la segunda es la advección, que implica el transporte de calor por medio de un transporte de masa, es decir el calor acarreado por el agua misma intercambiada. Esto se puede expresar como:

$$\frac{1}{\tau} \iiint_{\text{volumen}} \rho c_p (T_\tau - T_0) dV = \frac{1}{\tau} \int_0^\tau \iint_{\text{superficie del mar}} (Q_t - rQ_t - EL_t - Q_b - Q_c) dS d\tau + \frac{1}{\tau} \int_0^\tau \iint_{\text{frontera de las paredes}} \rho c_p TV_{\text{normal}} dP d\tau$$

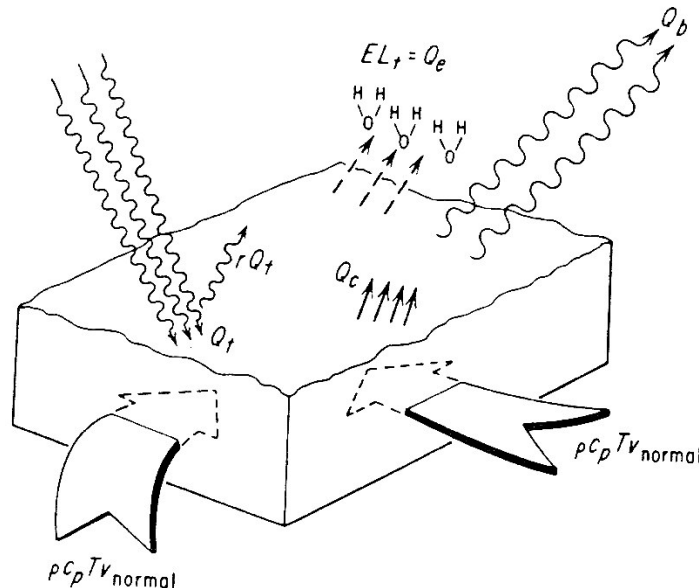


Figura B4. Balance de energía en un volumen unitario en la superficie del mar (Neumann y Pierson, 1966).

Donde:

ρ es la densidad

S es la superficie

V es el volumen considerado

P es el área de las paredes del volumen

τ es el tiempo final

T_0 es la temperatura inicial

T_τ es la temperatura final

Q_t es la radiación de onda corta que incide sobre la superficie del mar

r es el albedo (% de reflexión) de radiación de onda corta

E es la cantidad de agua evaporada de la superficie en $\text{gr cm}^{-2} \text{seg}^{-1}$

L_t es el calor latente a temperatura t , en cal/cm . ($L_t = 596 - 0.52t$)

Q_b es la diferencia entre la radiación de onda larga emitida por la superficie del mar y aquella que penetra al mar.

Q_c es la energía sensible intercambiada entre la atmósfera y el océano por convección y conducción directa.

V_{normal} es la cantidad de agua que entra al volumen por las paredes del volumen en un tiempo t .

$Q_e = EL_t$ es el calor perdido desde la superficie del mar debido a evaporación.

$\rho c_p T V_{normal}$ es el calor transportado perpendicular a las paredes del volumen.

rQ_t es la radiación de onda corta reflejada desde la superficie del mar.

Tomando una media de los procesos anteriores durante un lapso prolongado, se observa que la diferencia entre las temperaturas inicial y final es aproximadamente cero, por lo que se deduce que existe un balance entre dichos procesos. A partir de esta conclusión, se puede calcular términos individuales de la ecuación.

3. La temperatura de la superficie del mar-TSM

Observaciones satelitales de la TSM o SST (Sea Surface Temperature) contribuyen a entender la variabilidad regional y global del cambio climático y permiten visualizar una amplia variedad de flujos oceánicos. Estas observaciones son importantes por las siguientes razones. Primero,

porque la capa superior de 3m en la columna de agua tiene casi la misma capacidad calorífica, la capa superior del océano regula el sistema climático global donde la TSM es proporcional al calor almacenado en la capa superior del océano (Gill, 1982). Segundo, las distribuciones espaciales y temporales de los flujos atmosféricos de vapor de agua y calor son funciones de la temperatura superficial. Tercero, los patrones de gradientes de temperatura superficial asociados con los sistemas de corrientes, remolinos y regiones de afloramiento hacen que estos procesos sean visibles en imágenes de TSM.

A partir de los 80's con el lanzamiento de satélites, ya van más de tres décadas de observaciones satelitales de la TSM. Estas contribuyen a estudiar el clima global en varios niveles, mediante investigaciones en oceanografía física, apoyo a la industria pesquera dando rutas de pesca a los buques, y pronóstico del tiempo. Ejemplos de estudios a largo plazo incluyen los cambios en los patrones de TSM asociados con variaciones interanuales del clima como los fenómenos La Niña y El Niño en el Pacífico ecuatorial. Ejemplos de aplicaciones de TSM a corto plazo incluyen la delineación de los frentes oceánicos, regiones de afloramiento y remolinos oceánicos. La identificación y seguimiento de estas características requiere una precisión de al menos $\pm 0.5K$ (Walton, 1998).

Dado que la TSM depende de las propiedades atmosféricas y de concurrentes medidas en la superficie, dos conjuntos de procesos deben ser considerados. El primero es el efecto del sol y la atmósfera en las radiancias recibidas, el segundo es el efecto de la superficie oceánica y procesos cerca de la superficie en mediciones de TSM.

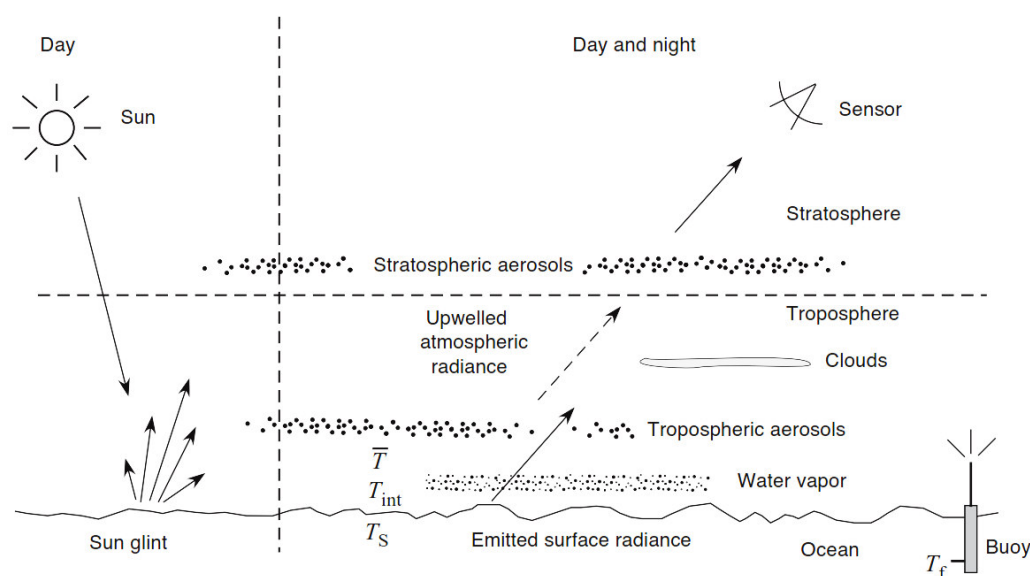


Figura B9. Esquema de las radiancias y propiedades atmosféricas y oceánicas involucradas en la estimación de la TSM (Seelye, 2014)

La figura B9 ilustra las radiancias y propiedades atmosféricas involucradas en la estimación de las radiancias superficiales del océano. En la figura, \bar{T} es la temperatura característica de la atmósfera inferior, T_f es la temperatura de la boya o temperatura de Bulk medida a 0.3-1m de profundidad, y T_s es la temperatura superficial (skin temperature), que es la temperatura en una delgada capa en la parte superior de la columna de agua. La diferencia entre T_s y la TSM es que T_s es la temperatura física de la capa más cercana a la superficie y que está en contacto directo con la atmósfera, mientras que TSM es una aproximación a T_s que depende de las técnicas de medición. Una variedad de procesos en la superficie y cerca de la superficie determinan la TSM y el perfil de temperatura en la capa superior del océano. Estos procesos incluyen el calentamiento solar durante el día y el enfriamiento radiativo en la noche, enfriamiento evaporativo, y la mezcla de viento y olas (figura B10). En general, la capa superior del océano se calienta durante el día por acción del sol y se enfría durante la noche. Debido a estos procesos, la diferencia $\Delta T = T_s - T_f$ entre la temperatura de superficie y la temperatura de bulk puede variar como mucho $\pm 1\text{K}$ (Katsaros, 1980).

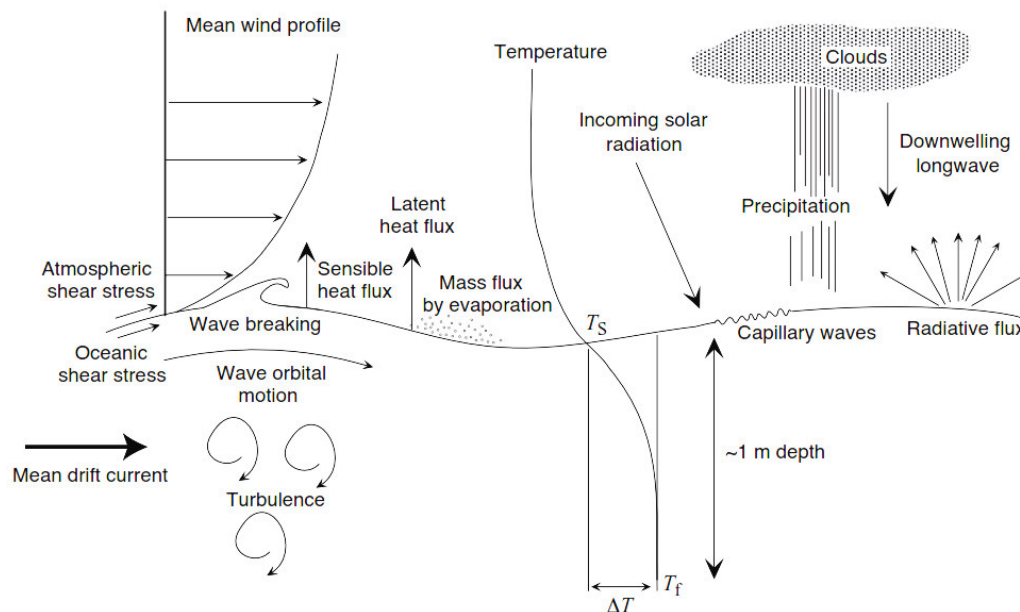


Figura B10. Los factores que determinan la temperatura en la superficie y las temperaturas al interior de la capa del océano, donde ΔT es la diferencia entre la temperatura de superficie y la temperatura de bulk (Seelye, 2014).

Debido a que diferentes instrumentos miden la TSM de diferentes maneras, el término TSM se refiere a las temperaturas medidas sobre diferentes intervalos de profundidad (figura B11). La primera es la temperatura de superficie T_s (skin temperature), que

contribuye a la determinación de los flujos de calor y humedad en el océano y atmósfera. Debido al enfriamiento evaporativo y radiativo, T_S es generalmente más baja que la temperatura interna. La segunda es la temperatura medida por un radiómetro infrarrojo operando en $11\mu\text{m}$, mostrada en la figura como $T_{11\mu\text{m}}$, que es la temperatura radiativa emitida por encima de los $30\mu\text{m}$ de la columna de agua. En general, $T_{11\mu\text{m}}$ es ligeramente superior a T_S , y dada la actual tecnología, esta tiene una medida aproximada a T_S . Una temperatura similar puede ser definida en $4\mu\text{m}$. La tercera es la temperatura $T_{10\text{GHz}}$, que es medida a 10GHz por un radiómetro en el rango de los microondas, esta es la temperatura radiativa emitida por encima de los $1\text{-}2\text{ mm}$ de la columna de agua. La última es la temperatura de bulk T_f que es medida a una profundidad de $0.3\text{-}1\text{ m}$ por una boya.

Los términos T_S y $T_{11\mu\text{m}}$ responden a los cambios en el balance evaporativo, conductivo y radiativo dentro de unos pocos segundos; $T_{10\text{GHz}}$ es dominado por procesos de viscosidad y responde en escalas de tiempo del orden de los minutos. Para vientos débiles, T_f responde a calentamientos y enfriamientos diurnos en el orden de las horas; para vientos fuertes, T_f responde a periodos más cortos.

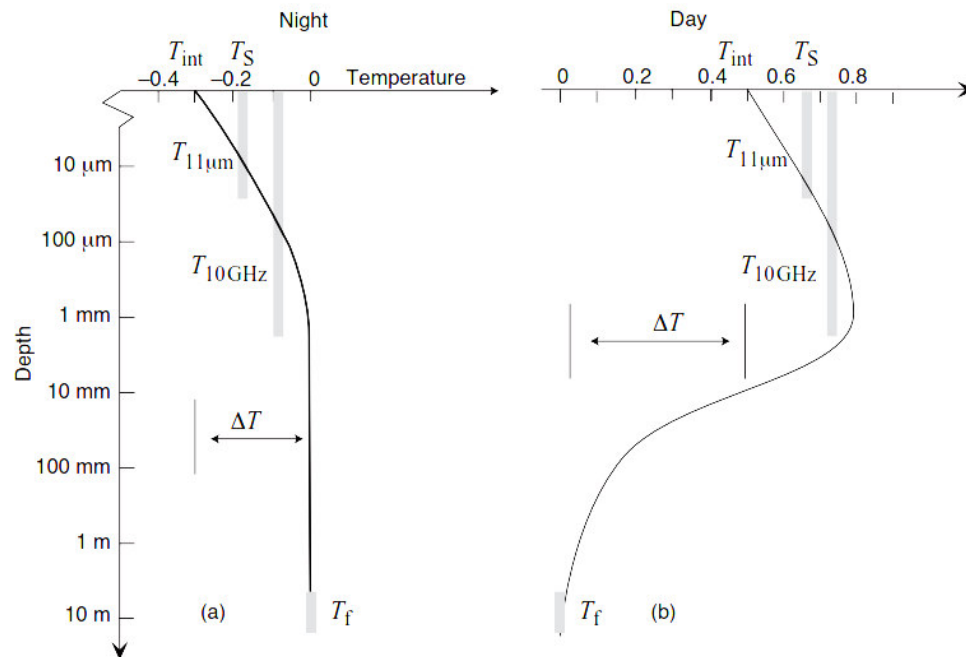


Figura B11. Comparación de los perfiles de temperatura para el día y la noche. La profundidad está en una escala logarítmica. (a) Caso de vientos fuertes en el día y la noche, donde $U > 6\text{ ms}^{-1}$. (b) Caso de vientos débiles e insolación solar durante el día que producen una capa superior del océano estratificada (Seelye, 2014).

4. Oscilación Sur

En el siglo XIX, los científicos ingleses querían encontrar una forma de predecir la ocurrencia o no de lluvias en la India en los meses usualmente lluviosos (junio, julio, agosto) porque había años que no ocurrían lluvias y generaban hambruna en la India. Durante el invierno del HN (Hemisferio Norte) durante enero hay menor radiación sobre la India, por lo tanto el continente es más frío que el océano, el cual genera un gradiente de presión entre estas dos áreas y consecuentemente se originan vientos del noreste (NE) dirigiéndose del continente hacia el océano (figura B12a). Durante el periodo con vientos del NE, una alta presión domina sobre la India; y no se producen precipitaciones. En verano del HN durante julio (figura B12b) la dirección de los vientos cambia totalmente; ahora el viento se dirige del océano llevando aire húmedo al continente. Durante este periodo hay abundante precipitación sobre el continente. Al cambio brusco en la dirección de los vientos de una estación a otra, suele llamarse “**monzón**”. Durante el verano del HN el monzón del suroeste trae mucha lluvia a la India, por lo que se llama también monzón lluvioso. Pero hay años que no hay este tipo de monzón y trae como consecuencia la sequía en la India.

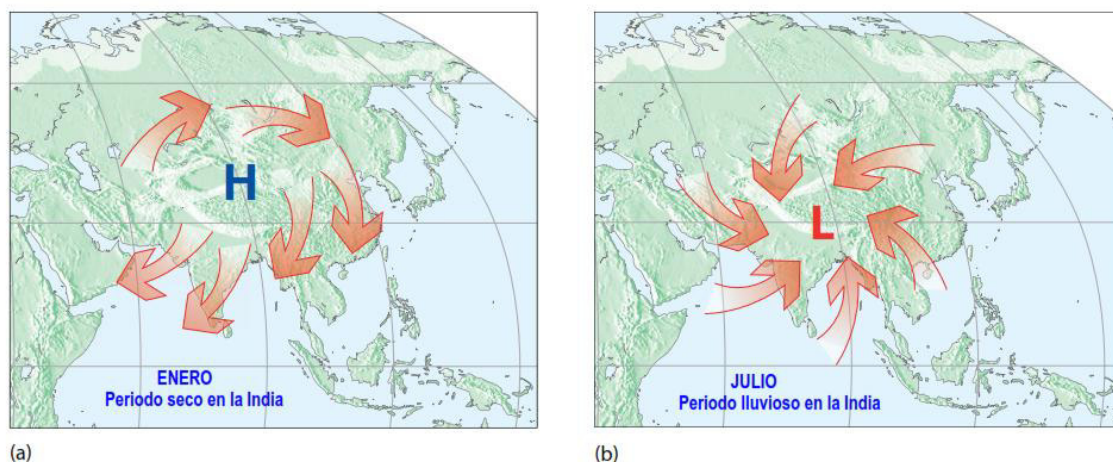


Figura B12. (a) Monzón del noreste y (b) Monzón del suroeste (Adaptado de Rohli and Vega 2018).

En aquella época los científicos intentaban explicar el por qué ocurría la interrupción del periodo lluvioso. Interesados en el tema, los ingleses crearon en 1875 el Departamento de Meteorología en la India con el propósito de encontrar una forma de predecir los monzones. Hubo esfuerzos de muchos científicos de asociar la fluctuación de los monzones con las manchas solares, otros trataron de encontrar alguna relación con las

inundaciones del río Nilo y con la nieve en el Himalaya. Pero parece que la nueva hambruna que sacudió en 1877 la India, dio pistas a los científicos para encontrar el origen de su variabilidad. Ellos quedaron sorprendidos al observar una alta presión, muy intensa sobre la India entre 1876 y 1878 con pico entre julio y setiembre de 1877. Y se encontraron con otra sorpresa, que la alta presión se extendía hasta Australia y la China, y en forma contraria en América del Sur se había formado una baja presión (figura B13). Es decir, habían encontrado que la variabilidad interanual del monzón era un fenómeno de escala global.

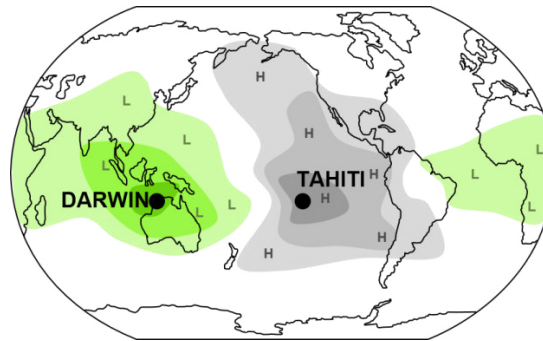


Figura B13. Patrón de anomalías de presión asociado a una fase positiva del IOS. Este se asocia a presiones sobre lo normal en el Pacífico oriental, y debajo de lo normal en Indonesia y Australia. Fuente: WPC International Desks.

Entre 1899 y 1900 nuevamente hubo hambruna en la India por falta de lluvias. En 1904, el matemático Gilbert Walker es nombrado Director General del Observatorio de la India y de inmediato recluta jóvenes para ayudar en los cálculos estadísticos, y tratar de encontrar una forma objetiva de pronosticar los monzones en función de los cambios que ocurren a escala global, ejemplo las variaciones de la presión y precipitación que ocurren en América del Sur, África y Australia.

El Dr. Walker fue director del observatorio por 20 años y durante ese tiempo se dedicó a investigar sobre la relación que existía entre la alta presión sobre la India y baja presión en Sudamérica y publicó numerosos trabajos sobre la forma de pronosticar el monzón. Walker observó que la alta presión entre Australia y la India cambiaba a baja presión en otros años y favorecían la ocurrencia de lluvias en la India, y nuevamente en los siguientes años volvía a alta presión, en forma análoga la baja presión al frente de Chile (Isla de Pascua) cambiaba a alta presión y luego, nuevamente volvía a baja presión, pero siempre opuesta a la de Australia. Este fenómeno ocurría cada 3 a 4 años. A esta

oscilación de la presión sobre el Océano Pacífico Sur con dos centros de acción (una alta y otra baja), una en Australia y otra al frente de Chile, Walker lo llamó “Oscilación Sur”. La palabra “Sur” fue usada porque estos centros de acción están ubicados en el Hemisferio Sur. A la diferencia de presión entre Tahití (o Pascua) y Darwin se denomina Índice de Oscilación Sur o IOS (Figueroa, 2001).

Los científicos ahora reconocen que El Niño y la Oscilación Sur son simplemente aspectos del mismo modo acoplado del sistema océano-atmósfera. La evidencia observacional para esta conexión es aparente en la figura B14, que muestra la correlación inversa de las variaciones de la TSM y el gradiente de presión superficial (Neelin y Latif, 1998).

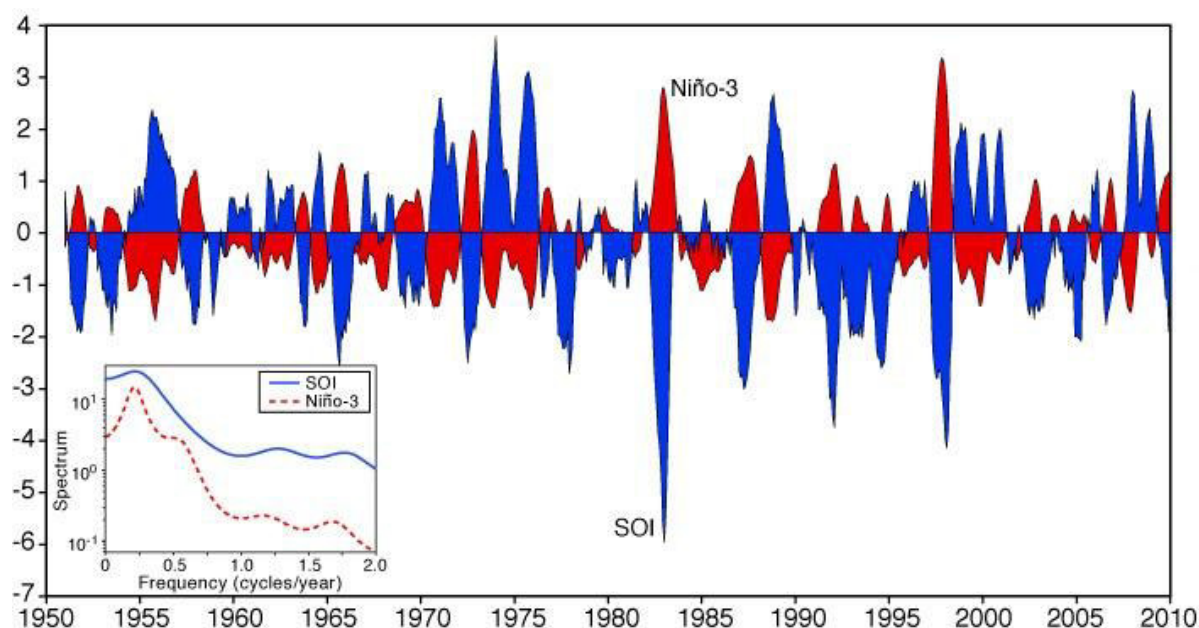


Figura B14: Covariabilidad entre la ATSM y el índice IOS. En este gráfico la parte roja representa las anomalías de la temperatura superficial del mar, promediado sobre el Pacífico central-este en el Ecuador para el periodo 1950-2010. La parte azul representa el índice de Oscilación Sur (Neelin y Latif, 1998).

5. Fenómeno El Niño y La Niña

El Niño es un fenómeno que resulta de la interacción océano-atmósfera y ocurre en el Pacífico Ecuatorial en la escala de tiempo 2 a 7 años. El cual consiste en el calentamiento anormal de las aguas superficiales del Pacífico Ecuatorial Central y Este, cuyas consecuencias inmediatas son la excesiva precipitación en Ecuador y costa norte del Perú, disminución de la producción pesquera en Ecuador y Perú, y el calentamiento anormal del aire en todo el territorio peruano, en especial en la costa peruana.

La niña es un fenómeno opuesto a El Niño. El cual consiste en el enfriamiento anormal de las aguas superficiales del Pacífico Ecuatorial Central y Este, cuyas consecuencias inmediatas son la disminución o ausencia de la precipitación en la costa norte del Perú, aumento de la producción pesquera en Ecuador y Perú, y enfriamiento anormal del aire en todo el territorio peruano.

El primer científico que encontró que existe una estrecha relación entre la Oscilación Sur y El Niño fue Jacob Bjerknes (1969). Antes se creía que El Niño era un fenómeno netamente oceánico, y la Oscilación Sur netamente atmosférico. Actualmente se sabe que entre El Niño- La Niña y la Oscilación Sur hay una estrecha relación, por lo que algunos científicos prefieren usar la palabra ENSO (o ENOS) para referirse a las variaciones atmosféricas y oceánicas relacionados al Niño y la Oscilación Sur.

Otros científicos creen que la Oscilación Sur no es simplemente la variación de la presión atmosférica, es un proceso de retroalimentación entre el océano y la atmósfera, el cual incluye la variación interanual de la temperatura de la superficie del mar y la presión atmosférica en el Pacífico Tropical. Por lo que sugieren que El Niño y La Niña son las fases extremas de la Oscilación Sur. A continuación se describen algunos aspectos de la relación entre el viento, IOS y la temperatura de la superficie del mar durante El Niño y La Niña.

- a) El IOS positivo ocurre cuando hay alta presión al frente de Chile (ejemplo, Isla de Pascua) y baja presión en Australia (Darwin); y el IOS negativo ocurre en caso contrario.
- b) El Niño está asociado a grandes valores de IOS negativo, vientos alisios débiles y a un calentamiento del Océano Pacífico Ecuatorial Central y Este (3°C a 4°C).

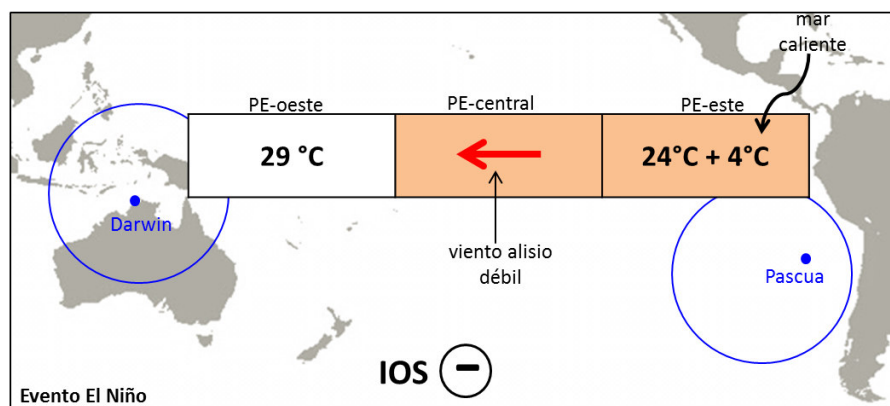


Figura B15. Fase negativa del IOS

- c) La Niña está asociada a grandes valores de IOS positivo, vientos alisios intensos y un enfriamiento del océano PE Central y Este (Figura B16).

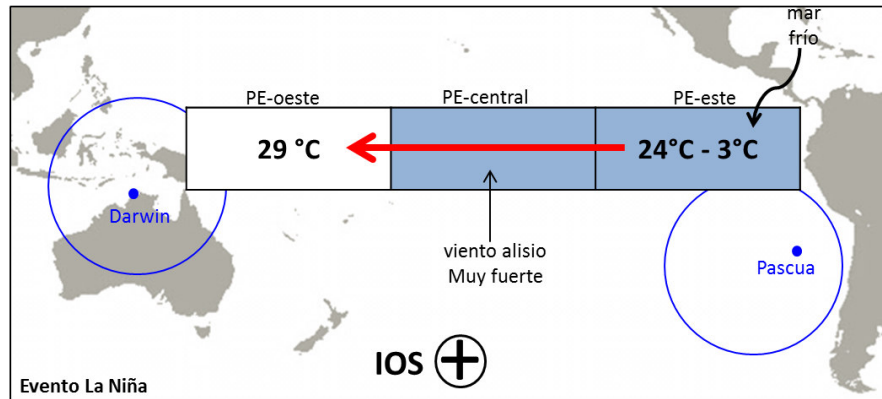


Figura B16. Fase positiva del IO

- d) Los vientos alisios intensos en el PE durante La Niña se deben al fuerte gradiente de presión entre la región PE-oeste y PE-este. Este fuerte gradiente de presión se debe al gradiente de temperatura de la superficie del mar entre estas dos zonas.

Los vientos se generan debido al gradiente de presión. Por ejemplo vientos alisios en el Océano Pacífico Ecuatorial dependen de la diferencia de presión entre la región Este y Oeste del Pacífico Ecuatorial. Por lo tanto, queda muy claro que la intensificación o debilitamiento de los vientos alisios en el Pacífico Ecuatorial se debe al gradiente de presión entre la región este y oeste del Pacífico. Esto es muy importante para entender El Niño y La Niña. Pero ¿Por qué hay alta o baja presión? La respuesta es simple, porque el mar es frío o caliente, en otras palabras depende de la TSM. Si el mar es caliente (Figura B17) hay fuente de calor (Q) y evaporación desde la superficie del mar.

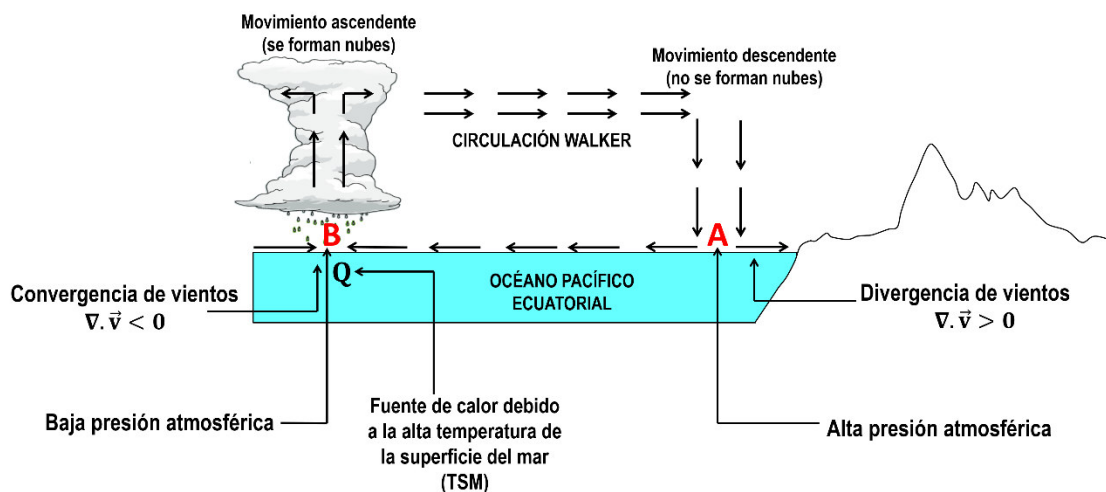


Figura B17. Interacción entre la temperatura del mar y los vientos.

Al existir evaporación, el aire se humedece y los vientos convergen hacia esta zona y favorece la formación de nubes de desarrollo vertical. Cuando se forman las nubes y luego precipita, la convergencia se intensifica. Esta convergencia a su vez ayuda a crecer a las nubes y el desarrollo de éstas nuevamente aumenta la convergencia de los vientos y así sucesivamente. Es decir se genera un proceso de retroalimentación entre la convección y los vientos convergentes. Por conservación de masa el aire no puede acumularse en la zona de convergencia, tiene que existir ascenso del aire, con el cual disminuye el peso del aire sobre el mar caliente, por lo tanto hay baja presión (B). Por otro lado, en la región de mar frío, el aire adyacente es frío y seco. El aire es más denso y consecuentemente hay descenso del aire y con ello se genera la divergencia de los vientos, el aire pesa más, en otras palabras hay alta presión (A).

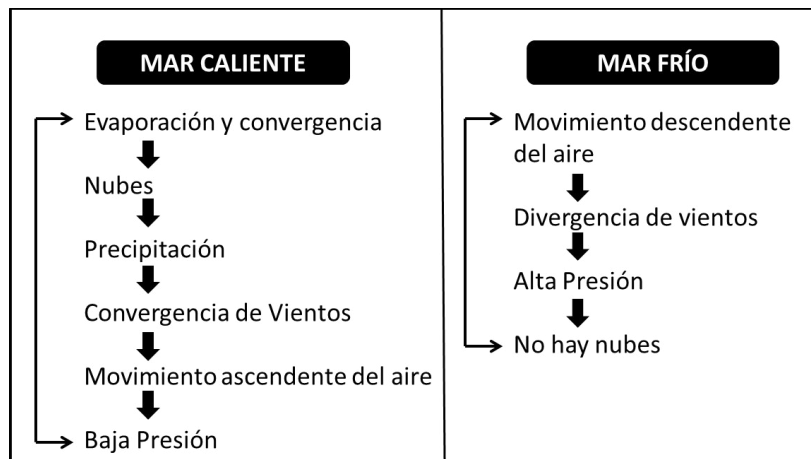


Figura B18. Resumen de los procesos de interacción océano-atmósfera

La radiación solar calienta el Pacífico Ecuatorial en forma homogénea, sin embargo la TSM en el PE no es uniforme, ésta es debido a los vientos. Pero alguien se preguntará, quién es primero: los vientos que cambian la temperatura de la superficie del mar, o la temperatura del mar, que cambia los vientos. No hay respuesta a esta pregunta. La atmósfera y el océano es un solo sistema, ninguno de ellos está aislado. Si no existiera el océano, la atmósfera sería totalmente distinta a la que conocemos, de igual forma, si no existiera atmósfera, el océano sería muy diferente. El clima que conocemos es producto de la interacción océano-atmósfera (Figueroa, 2001).

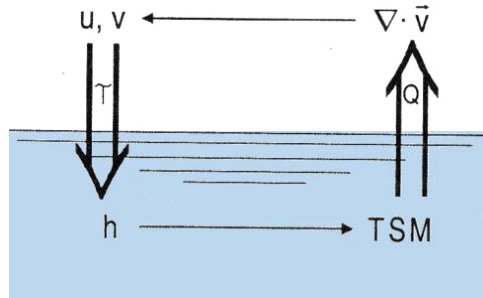


Figura B19. En esta figura se puede observar que el viento hace variar la profundidad de la termoclina (h) a través de la fricción del viento (τ) y ésta hace variar la TSM. Pero también (h) puede variar debido a la propagación de las ondas del océano.

Durante El Niño el PE oriental se calienta y durante La Niña se enfría (figura B19). ¿Por qué ocurre este calentamiento o enfriamiento? Porque las ondas Kelvin al propagarse hacia el este profundizan la termoclina (durante El Niño) o reducen la profundidad de la termoclina (durante La Niña).

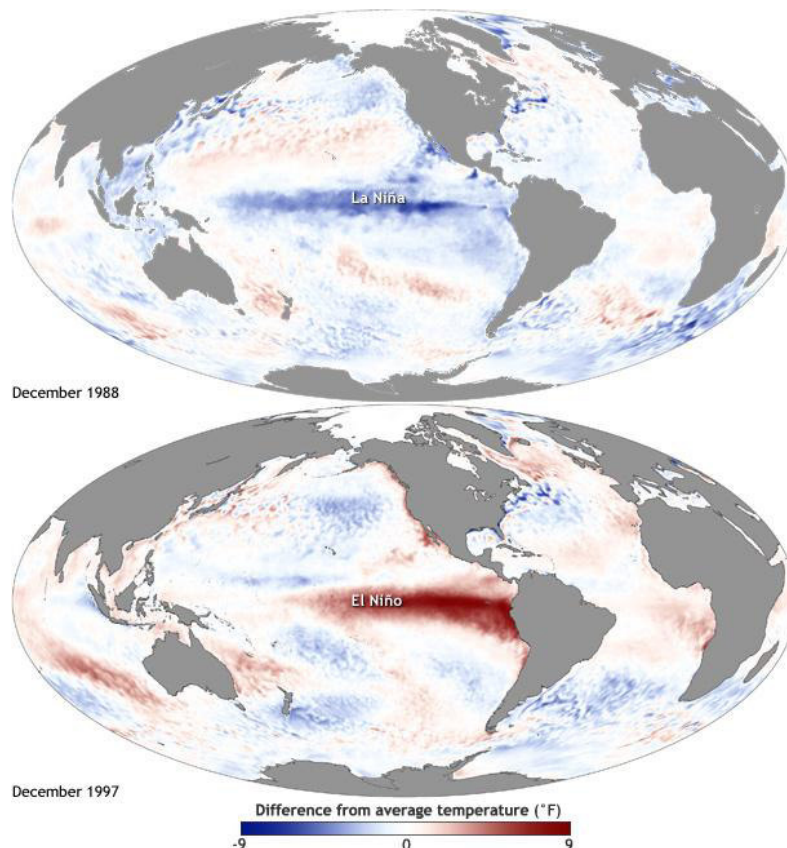


Figura B20. Mapas de anomalías de la temperatura superficial del mar durante un evento fuerte La Niña (parte superior) y un evento fuerte El Niño (parte inferior). Imagen extraída del portal www.climate.gov de la NOAA.

Cuando los vientos alisios se debilitan el nivel del mar en la parte oeste baja unos 20 cm y la profundidad de la termoclina se reduce unos 20 cm. Los cambios más dramáticos ocurren en la parte oriental del oeste frente a la costa peruana-ecuatoriana. El Nivel del mar sube unos 15 cm, pero la profundidad de la termoclina aumenta hasta 150m.

Por otro lado, cuando los vientos alisios son muy intensos en el Pacífico Ecuatorial, la altura del nivel del mar en la región PE-oeste aumenta unos 10cm y la profundidad de la termoclina se hunde unos 20 m respecto a su valor normal; y en la parte PE-este (frente al Perú) la altura del nivel del mar baja unos 5cm, pero la profundidad de la termoclina se reduce aproximadamente a 30m (figura B21 y B22). En la tabla B1 se resumen las principales características de los eventos El Niño y La Niña.

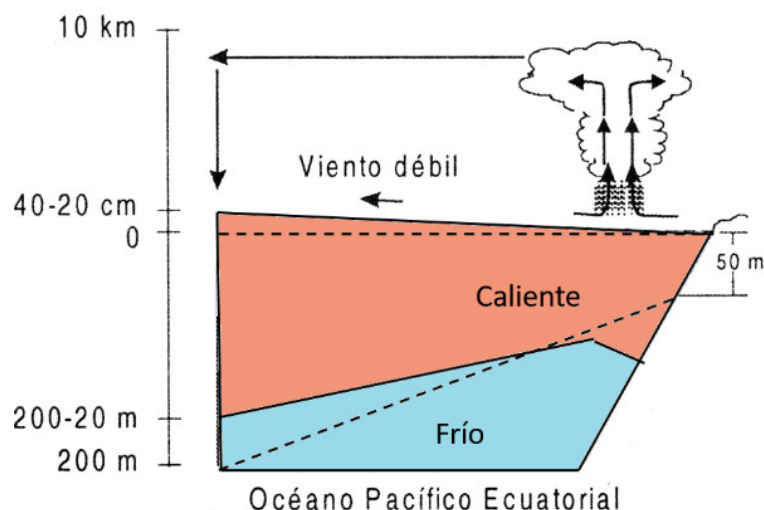


Figura B21. Profundidad de la termoclina durante El Niño (adaptado de Figueroa, 2001)

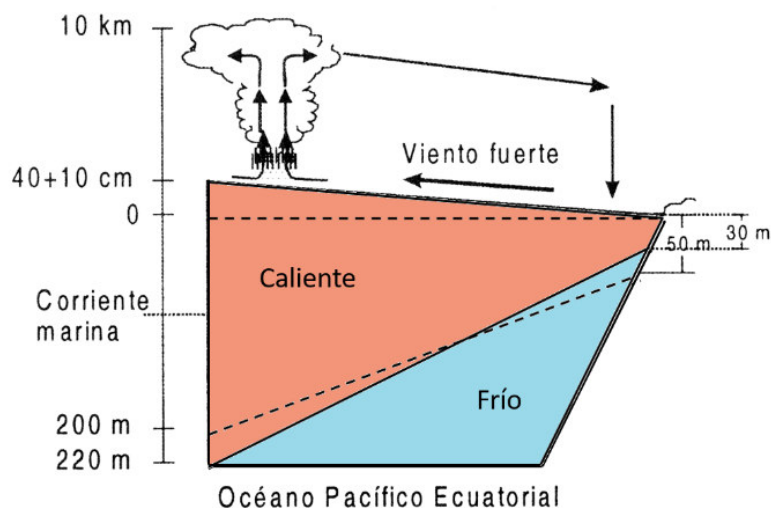


Figura B22. Profundidad de la termoclina durante La Niña (adaptado de Figueroa, 2001)

	EL NIÑO	LA NIÑA
TSM en el Pacífico Ecuatorial Central y Este	Sube	Baja
Vientos alisios sobre el Océano Pacífico Ecuatorial (PE)	Débiles	Intensos
Presión atmosférica en Darwin (al oeste del Pacífico Sur)	Alta	Baja
Presión atmosférica en Pascua (al este del Pacífico Sur)	Baja	Alta
Profundidad de la termoclina en la región PE-oeste	Se reduce	Se profundiza
Profundidad de la termoclina en la región PE-este	Se profundiza	Se reduce
IOS	(-)	(+)
Otros nombres	<ul style="list-style-type: none"> - Fase cálida ENSO - Evento cálido - Fase cálida de OS 	<ul style="list-style-type: none"> - Fase fría ENSO - Evento frío - Fase fría de OS

Tabla B1. Principales características de los eventos El Niño y La Niña

En la figura B23 y B24 se muestran en forma esquemática el efecto de El Niño y La Niña sobre el Perú. Es importante mencionar que La Niña es más próxima a un año normal.

Circulación Walker en año Normal

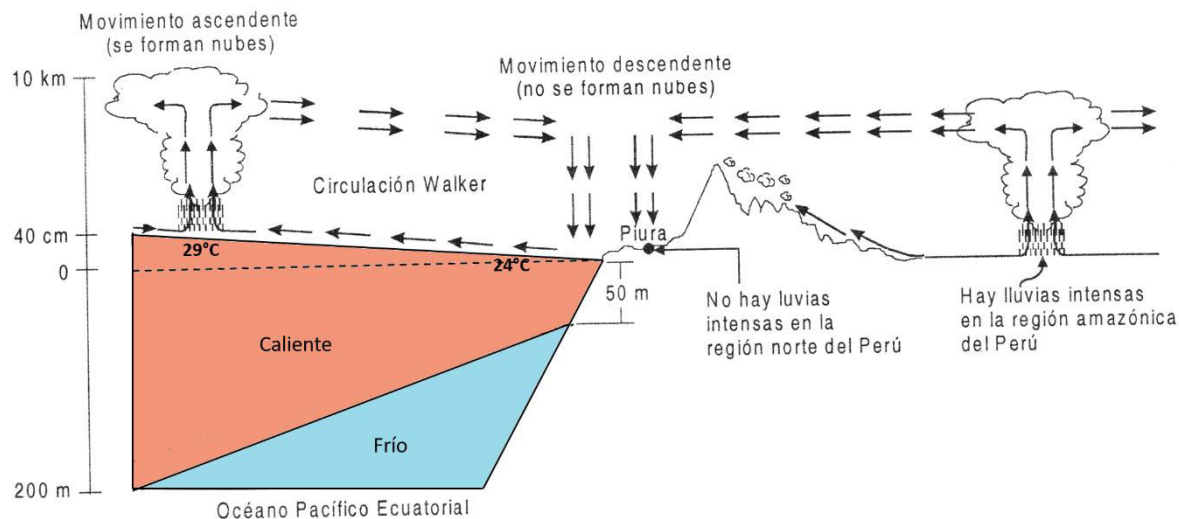


Figura B23. Efecto de un evento El Niño sobre Perú

Circulación Walker en año Niño

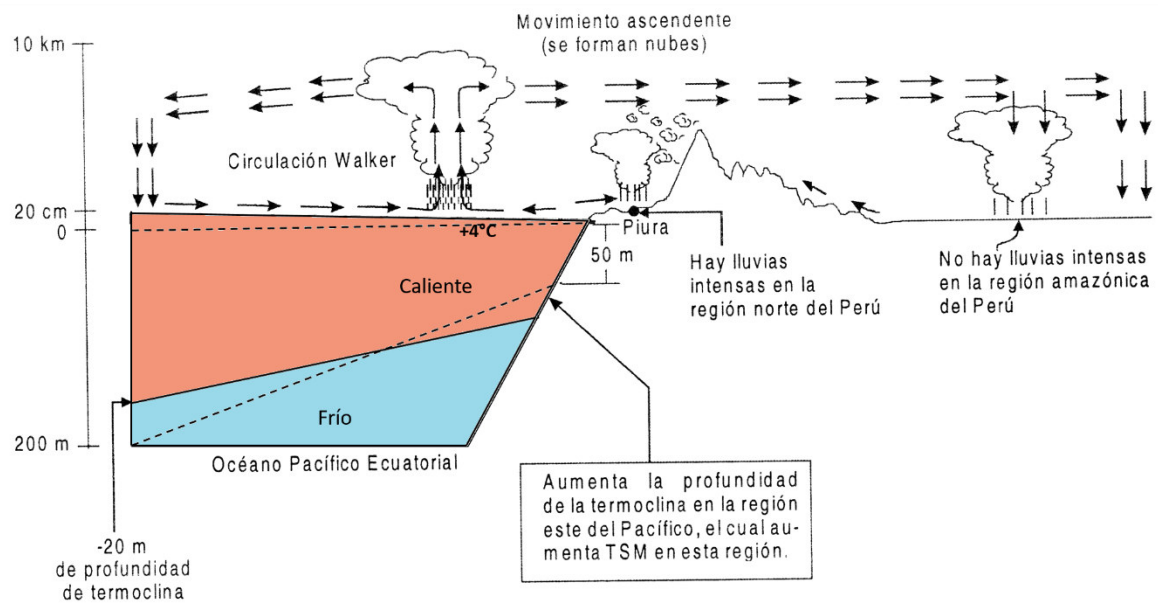


Figura B24. Efecto de un evento La Niña sobre Perú.

ANEXO 3: Funciones Ortogonales Empíricas en IDL

1) Aplicación del método SVD en IDL

F= Matriz de datos centrados =

-0.32500002	0.037499994	0.27500004	0.10000005	-0.70000005	-0.38749993
-0.12500000	0.33750001	0.27500004	0.20000005	-1.0000000	1.0125000
-0.12500000	-0.36250001	-0.12499999	5.9604645e-008	0.099999994	-1.4875000
0.17500001	0.63750005	-0.12499999	-0.099999964	-0.60000002	1.4125000
0.27500001	-0.062500007	-0.72500002	0.20000005	0.099999994	0.71250004
0.47499999	0.037499994	0.77499998	-0.099999964	0.70000005	1.3125000
0.17500001	0.23750000	-0.22499999	-0.19999993	1.3000001	-1.7874999
-0.52499998	-0.86249995	-0.12499999	-0.099999964	0.099999994	-0.7874999

Donde:

N=8 (# de muestras)

P=6 (# de variables)

$$A = U W V^T$$

a) Aplicación del método a la matriz de covarianza

A=(1./((N-1)) * (transpose(Double(F)) ## double(F)))

la_svd, A, W, U, V

autovalores = W*(N-1)

12.221240	2.5048764	1.2895764	1.0946537	0.11639225	0.030761923
-----------	-----------	-----------	-----------	------------	-------------

porcentaje de la varianza=W / Total(W) * 100.0

68.655015	18.200058	5.5964888	5.6842980	0.60439954	0.15973995
-----------	-----------	-----------	-----------	------------	------------

AUTOVECTORES=U

-0.088657213	0.39596897	0.18090864	-0.14955549	0.60026311	-0.64802615
-0.17932615	0.12990386	0.35803141	-0.85696638	-0.17116517	0.24308790
-0.10140010	0.063783161	-0.91073258	-0.37422604	0.12713507	0.0027293033
-0.045709404	-0.10965375	0.079882924	0.058835612	0.75018134	0.64286216
0.32404258	0.86728438	-0.056191942	0.15518403	-0.12915719	0.31447259
-0.91793633	0.24095461	-0.010626947	0.27505078	-0.12153081	0.093798537

AUTOVECTORES=V

-0.088657213	0.39596897	0.18090864	-0.14955549	0.60026311	-0.64802615
-0.17932615	0.12990386	0.35803141	-0.85696638	-0.17116517	0.24308790
-0.10140010	0.063783161	-0.91073258	-0.37422604	0.12713507	0.0027293033
-0.045709404	-0.10965375	0.079882924	0.058835612	0.75018134	0.64286216
0.32404258	0.86728438	-0.056191942	0.15518403	-0.12915719	0.31447259
-0.91793633	0.24095461	-0.010626947	0.27505078	-0.12153081	0.093798537

Cuando el método svd se aplica a la matriz de covarianza U=V(autovectores)

b) Aplicación del método a la matriz F^T

A=transpose (Double(F))

la_svd, A, W, U, V

autovalores = (W^2.)

12.221239	2.5048762	1.2895763	1.0946537	0.11639224	0.030761921
-----------	-----------	-----------	-----------	------------	-------------

porcentaje de la varianza== (W^2.) / Total(W^2.) * 100.0

68.655015	18.200058	5.5964888	5.6842980	0.60439954	0.15973995
-----------	-----------	-----------	-----------	------------	------------

AUTOVECTORES=U

0.088657213	0.39596897	-0.18090864	0.14955549	-0.60026311	0.64802615
0.17932615	0.12990386	-0.35803141	0.85696638	0.17116517	-0.24308790

0.10140010	0.063783161	0.91073258	0.37422604	-0.12713507	-0.0027293033
0.045709404	-0.10965375	-0.079882924	-0.058835612	-0.75018134	-0.64286216
-0.32404258	0.86728438	0.056191942	-0.15518403	0.12915719	-0.31447259
0.91793633	0.24095461	0.010626947	-0.27505078	0.12153081	-0.093798537

c) Aplicación del método a la matriz F

A=Double(F)

la_svd, A, W, U, V

autovalores = (W^2.)

12.221239	2.5048762	1.2895763	1.0946537	0.11639224	0.030761921
-----------	-----------	-----------	-----------	------------	-------------

porcentaje de la varianza== (W^2.) / Total(W^2.) * 100.0

68.655015	18.200058	5.5964888	5.6842980	0.60439954	0.15973995
-----------	-----------	-----------	-----------	------------	------------

AUTOVECTORES=V

0.088657213	0.39596897	-0.18090864	0.14955549	-0.60026311	0.64802615
0.17932615	0.12990386	-0.35803141	0.85696638	0.17116517	-0.24308790
0.10140010	0.063783161	0.91073258	0.37422604	-0.12713507	-0.0027293033
0.045709404	-0.10965375	-0.079882924	-0.058835612	-0.75018134	-0.64286216
-0.32404258	0.86728438	0.056191942	-0.15518403	0.12915719	-0.31447259
0.91793633	0.24095461	0.010626947	-0.27505078	0.12153081	-0.093798537

d) Aplicación del método a la matriz $\frac{F}{\sqrt{N-1}}$

A=(1/SQRT(N-1))*(Double(f))

la_svd, A, W, U, V

autovalores = (W^2.)*(N-1)

12.221239	2.5048762	1.2895763	1.0946537	0.11639224	0.030761921
-----------	-----------	-----------	-----------	------------	-------------

porcentaje de la varianza== (W^2.) / Total(W^2.) * 100.0

68.655015	18.200058	5.5964888	5.6842980	0.60439954	0.15973995
-----------	-----------	-----------	-----------	------------	------------

AUTOVECTORES=V

0.088657213	0.39596897	-0.18090864	0.14955549	-0.60026311	0.64802615
0.17932615	0.12990386	-0.35803141	0.85696638	0.17116517	-0.24308790
0.10140010	0.063783161	0.91073258	0.37422604	-0.12713507	-0.0027293033
0.045709404	-0.10965375	-0.079882924	-0.058835612	-0.75018134	-0.64286216
-0.32404258	0.86728438	0.056191942	-0.15518403	0.12915719	-0.31447259
0.91793633	0.24095461	0.010626947	-0.27505078	0.12153081	-0.093798537

2) Código para estimar el índice CP en IDL

```

pro INDICE_CP_KAO
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
data=fltarr(4,732)
OPENR, lun, 'C:/nino.ascii',/GET_LUN
readf, lun, data
close, lun
FREE_LUN, lun
an_sst_nino12=data(0,*)
an_sst_nino3=data(1,*)
an_sst_nino4=data(2,*)
an_sst_nino34=data(3,*)

anomalia_01=an_sst_nino12
anomalia_02=an_sst_nino3
anomalia_03=an_sst_nino4
anomalia_04=an_sst_nino34

```



```

;;;;;;;;;;;;;
ncdf = Obj_New('ncdf_data', 'C:\HadISST_sst.nc')
time = ncdf -> ReadVariable('time')
lat = ncdf -> ReadVariable('lat')
lon = ncdf -> ReadVariable('lon')
; time_bnds = ncdf -> ReadVariable('time_bnds')
tsm = ncdf -> ReadVariable('sst')
Obj_Destroy, ncdf

jd = (time/24.D)+Julday(1, 1, 1, 0, 0, 0)
Caldat, jd, month, day, year

;Considerando el periodo 1950-2010
a_inicial=1950
a_final=2010
bsse_inicial=1982
base_final=2010

bad=where(year eq a_inicial)
bad3=where(year eq bsse_inicial)
bad4=where(year eq base_final+1)

sst=tsm[:,*,bad[0]:bad[0]+(a_final-a_inicial+1)*12-1]

;extrayendo el periodo base
sst_base=tsm[:,*,bad3[0]:bad4[0]-1]

lon_anom_360 = Rebin(lon, n_elements(lon), n_elements(lat))
lat2=Reform(lat, 1, n_elements(lat))
lat_anoma = Rebin(lat2, n_elements(lon), n_elements(lat))
lon_ANOMA = ((lon_ANOM_360 + 180) MOD 360) - 180

s=size(SST)
SST=SHIFT(SST, s(1)/2 )
sst_base=SHIFT(sst_base, s(1)/2 )
lon_ANOMA =SHIFT(lon_ANOMA , s(1)/2 )
lat_ANOMA =SHIFT(lat_ANOMA , s(1)/2 )

lat_anoma=reverse(lat_anoma,2)
sst=reverse(SST,2)
sst_base=reverse(sst_base,2)

sur=-30
norte=30
oeste=140
este=-70

near = Min(Abs(lat_anoma[0,*]- sur),index)
lat_min_index=index
near = Min(Abs(lon_ANOMA[:,0]- oeste),index)
lon_min_index=index
near = Min(Abs(lat_anoma[0,*]- norte),index)
lat_max_index=index

```

```

near = Min(Abs(lon_ANOMA[*,0]- este),index)
lon_max_index=index

sst =sst[lon_min_index:lon_max_index, lat_min_index:lat_max_index, *]
sst_base=sst_base[lon_min_index:lon_max_index,
lat_min_index:lat_max_index, *]

sur2=-10
norte2=0
oeste2=-90
este2=-80

lat_sst = lat_anoma[lon_min_index:lon_max_index,
lat_min_index:lat_max_index]
lon_sst = lon_ANOMA[lon_min_index:lon_max_index,
lat_min_index:lat_max_index]

near = Min(Abs(lat_sst[0,*]- sur2),index)
lat_min_index2=index
near = Min(Abs(lon_sst[*,0]- oeste2),index)
lon_min_index2=index
near = Min(Abs(lat_sst[0,*]- norte2),index)
lat_max_index2=index

near = Min(Abs(lon_sst[*,0]- este2),index)
lon_max_index2=index

p=size(sst)
pp=size(sst_base)
; sst1=sst[*, *, 1]

lat_ncep=lat_sst
lon_ncep=lon_sst
sst_temp=sst_base

;Calculando las anomalías
sst_subset = reform(sst,p[1],p[2],12,p[3]/12.); full periodo
sst_subset2 = reform(sst_base,pp[1],pp[2],12,pp[3]/12.); periodo 1981-2010
monthly_means =
rebin(rebin(sst_subset2,pp[1],pp[2],12),p[1],p[2],12,p[3]/12.)
sst_anomalies= reform(sst_subset-monthly_means,p[1],p[2],p[3])
nulos=where(sst_anomalies lt -1000. or sst_anomalies gt 1000)
sst_anomalies(nulos)=!VALUES.F_NAN
ninol2=sst_anomalies[lon_min_index2:lon_max_index2,
lat_min_index2:lat_max_index2, *]

promedio=fltarr(732)

for x=0, 732-1 do begin
promedio[x]=mean(ninol2[*,*,x],/NAN)
endfor

promedio=reform(promedio,1,732)
nrot=2
sst_anomalies= reform(sst_anomalies,p[1]*p[2],p[3])
sst_anomalies3=sst_anomalies

```

```

amm=fltarr(p[1]*p[2],p[3])
for j=0,9211-1 do begin
for k=0,1-1 do begin
amm[j,k]=regress(promedio[k,*],reform(sst_anomalies(j,*)))
endfor
endfor

amm1=fltarr(p[1]*p[2],1)
amm1[:,0]=amm[:,0]

for op=0,p[3]-1 do begin
sst_anomalies3[:,op]=sst_anomalies[:,op]-amm1*promedio[0,op]
endfor

sst_anomalies= reform(sst_anomalies,p[1],p[2],p[3])
sst_anomalies3= reform(sst_anomalies3,p[1],p[2],p[3])

isize=size(sst_anomalies)
nx=isize(1)
ny=isize(2)
nz=isize(3)
print,nx,ny,nz
iposx=fltarr(nx*ny)
iposy=fltarr(nx*ny)

iloc=-1L

for ix=0L,nx-1L do begin
for iy=0L,ny-1L do begin
xcheck=fltarr(nz)
xcheck(*)=sst_anomalies(ix,iy,*)
whrgood=where(finite(xcheck) ne 0)
isize=size(whrgood)
if whrgood(0) ne -1 then begin
; print,'total good points = ',isize(1)
itimes=isize(1)
iloc=iloc+1

if iloc eq 0 then sst1d=fltarr(nx*ny,isize(1))
sst1d(iloc,*)=xcheck(whrgood)
iposx(iloc)=ix
iposy(iloc)=iy
endif
endfor
endfor

nlocgood=iloc
sst1dgood=fltarr(nlocgood+1,itimes)
sst1dgood(*,*)=sst1d(0:nlocgood,*)
sst1d=sst1dgood
sst_anomalies4=sst1d

isize=size(sst_anomalies3)
nx=isize(1)
ny=isize(2)

```

```

nz=isize(3)
print,nx,ny,nz
iposx=fltarr(nx*ny)
iposy=fltarr(nx*ny)

iloc=-1L

for ix=0L,nx-1L do begin
  for iy=0L,ny-1L do begin
    xcheck=fltarr(nz)
    xcheck(*)=sst_anomalies3(ix,iy,*)
    whrgood=where(finite(xcheck) ne 0)
    isize=size(whrgood)
    if whrgood(0) ne -1 then begin
      itimes=isize(1)
      iloc=iloc+1
      if iloc eq 0 then sst1d=fltarr(nx*ny,isize(1))
      sst1d(iloc,*)=xcheck(whrgood)
      iposx(iloc)=ix
      iposy(iloc)=iy
    endif
  endfor
endfor

nlocgood=iloc
sst1dgood=fltarr(nlocgood+1,itimes)
sst1dgood(*,*)=sst1d(0:nlocgood,*)
sst1d=sst1dgood

sst_anomalie=sst1d
p=size(sst_anomalie)

;quitando la tendencia lineal
TAM =SIZE(sst_anomalie)
pendiente=fltarr(tam[1])
intercepto=fltarr(tam[1])
ff=findgen(tam[2])
for k=0, tam[1]-1 do begin
  result=linfit(ff,sst_anomalie[k,*])
  pendiente[k]=result[1]
  intercepto[k]=result[0]
endfor
sst_anomalies2=FLTARR(tam[1],tam[2])
for k=0, tam[1]-1 do begin
  sst_anomalies2[k,]=sst_anomalie[k,]-pendiente[k]*ff-intercepto[k]
endfor
for z=0,p[1]-1 do sst_anomalies2[z,]=sst_anomalies2[z,]-
mean(sst_anomalies2[z,*])

;PASO2:calculando la matriz de covariancia
matrix = (1/float(p[2]-1))*double(sst_anomalies2) ##
Transpose(sst_anomalies2);

la_SVD, matrix, W, U, V

percent_variance = W / Total(W) * 100.0
cumulative_variance=total(percent_variance,/cumulative)

```

```

print,'percent_variance'
print,percent_variance[0:3]
print,'total(percent_variance)'
print,total(percent_variance[0:3])

      s = Size(sst_anomalies2, /DIMENSIONS)
      eof = FltArr(s[1], s[0])

      FOR j=0,s[1]-1 DO BEGIN
        t = Transpose(sst_anomalies2) ## u[j,*];transpose(data original)##
        eof[j,*] = t / SQRT(Total(t^2));
      ENDFOR

      pc = FltArr(s[1], s[1])
      FOR j=0,s[1]-1 DO pc[j,*] = sst_anomalies2 ## eof[j,*];solo los
      pc[0,*]=pc[0,*]/STDEV(PC[0,*])

      EOF=fltarr(p[1],p[2])
      for j=0,p[1]-1 do begin
        for k=0,1-1 do begin
          EOF[j,k]=regress(pc[k,*],reform(sst_anomalies2(j,*)))
        endfor
      endfor
      eof=Transpose(eof)
      ;

      eof2d=fltarr(s[1],nx,ny)
      for iloc=0,nlocgood do begin
        eof2d[,iposx(iloc),iposy(iloc)]=eof[,iloc]
      endfor

      ;;;;;;;;;;;;;;
      eof2d=reform(eof2d,s[1],nx*ny)
      eof=eof2d

      for i=0, 0 do begin
        sign=[1,-1,1,-1]
        p=size(sst)
        lx=p[1]
        ly=p[2]

        eof_tsm = FltArr(p[3],lx*ly)
        eof2=transpose(eof)
        eof_tsm=sign[i]*eof2[,i]
        eof_tsm2=eof_tsm
        ;eof_tsm3=fltarr(1,p[3])
        eof_tsm3=sign[i]*pc[i,*]

      WRITE_CSV, 'C:\indice_pca_kao_cp'+string(i+1,format="(i02)")+'.CSV',eof_tsm3
      theEOF=reform(eof_tsm2,lx,ly)
      tam=size(theEOF)
      num_rows=tam(2)
      num_cols=tam(1)

      ;ajustando los valores a la paleta

```

```

minran=-1.5
maxran=1.5
interval=(maxran-minran)/256.
vector1=total(((findgen(256)*0)+interval),/cumulative)-interval+minran
vector2=findgen(256)
result=linfit(vector1,vector2)

theEOF2=float(theEOF*result[1]+result[0])

PS_Start,'C:\indice__kao_cp.ps'

DEVICE, DECOMPOSED = 0
;TVlct, r, g, b
cgDisplay, /FREE, XSIZE=10*p[1], YSIZE=10*p[2]
DEVICE, SET_FONT='Helvetica' ,/TT_FONT

Erase, COLOR=cgColor('ivory')
cgLoadCT, 1, NCOLORS=128, CLIP=[55, 255]
cgLoadCT, 3, NCOLORS=127, CLIP=[55, 255], BOTTOM=128, /REVERSE
TVLCT, cgColor('ivory', /TRIPLE), 255

cgColorbar, RANGE=[-1.5, 1.5], DIVISIONS=10,/right, /vertical,
FORMAT='(F8.1)', $
POSITION=[0.92, 0.1, 0.935, 0.91], ANNOTATECOLOR='black', XMINOR=0,
charsize=0.7, $
FONT=0, NCOLORS=255

;;;;;;;;;;;;;
MAP_SET, /mercator, 0, -145, POSITION=[0.1, 0.1, 0.9, 0.9], $
LIMIT=[-30, 140, 30, 290], $;límites del mapa [Latmin, Lonmin, Latmax,
Lonmax]
/NOERASE, /NOBORDER
warp = Map_Patch(theEOF2, lon_ncep, lat_ncep, /TRIANGULATE, MISSING=255)
warp2 = Map_Patch(theEOF, lon_ncep, lat_ncep, /TRIANGULATE, MISSING=255)

valor = Indgen(255)+1
cgCONTOUR,warp,Levels=valor, C_Colors=valor,$
Position=[0.1, 0.1, 0.9, 0.9], noerase=1,c_labels=256,Font=1,
xStyle=1,YStyle=1,charsize=2.5,NLEVELS=5,/fill,$
axiscolor='white',YTICKFORMAT = '(A1)', xTICKFORMAT =
'(A1)',xticks=1,yticks=1
;;;;;;;;;;;;;
;
;AGREGANDO LINEAS DE CONTORNO
cont=(findgen(100)-50)*0.2+0.1
cgCONTOUR,theEOF,Levels=cont,Color='black',c_charsize=0.5,thick=0.4,
/OnImage,c_charthick=1.2

MAP_SET, /mercator, 0, -145, POSITION=[0.1, 0.1, 0.9, 0.9], $
LIMIT=[-30, 140, 30, 290], $;límites del mapa [Latmin, Lonmin, Latmax,
Lonmax]
/NOERASE, /NOBORDER
cgmap_continents, COLOR=cgColor('black'),/FILL_CONTINENTS,/HIRES

```

```

patrones=['CP']

cgText, 0.45, 0.93,font=1,CHARTHICK=2.5,CHARSIZE=1.5,/Normal, 'Patron '+patrones[i]
cgText, 0.92, 0.925,font=1,CHARTHICK=2.5,CHARSIZE=1,/Normal, + cgSymbol('deg', /PS) +'C

;agregando latitud y longitud
xrange = [0, 150];[lonmin, lonmax]
yrange = [30, -30];[latmax, latmin]
xticknames = ['140' + cgSymbol('deg', /PS) + 'E', '170' +
cgSymbol('deg', /PS) + 'E', '160' + cgSymbol('deg', /PS) + 'W', '130' +
cgSymbol('deg', /PS) + 'W', '100' + cgSymbol('deg', /PS) + 'W', '70' +
cgSymbol('deg', /PS) + 'W']
yticknames = ['30' + cgSymbol('deg', /PS) + 'S', '15' + cgSymbol('deg',
/PS) + 'S', '0' + cgSymbol('deg', /PS) + '', '15' + cgSymbol('deg', /PS) +
'N', '30' + cgSymbol('deg', /PS) + 'N']
  AXIS, XAXIS = 0, X RANGE = xrange, COLOR =cgColor('black'), /DEVICE,
/NOERASE, XTICKINTERVAL = 30, $
  XMINOR = 1, XSTYLE = 1, XTICKNAME = xticknames, FONT =
0,CHARSIZE=0.9
  AXIS, YAXIS = 0, Y RANGE = yrange, COLOR =cgColor('black'), /DEVICE,
/NOERASE, YTICKINTERVAL = 15, $
  YMINOR = 1, YSTYLE = 1, TICKLEN = 0.01, YTICKNAME =
yticknames, FONT =0,CHARSIZE=0.9
;ploteando la componente principal del modo 1

PS_End, /jpeg, RESIZE=200,density=500 ; idl 7.1

endfor

END

```

REFERENCIAS

- **Ashok, K., Behera, S., Rao S., Weng, H., Yamagata, T.** (2007). El Niño Modoki and its possible teleconnection. *J. Geophys. Res.*, 112, C11007.
- **Ashok, K., and Yamagata T.** (2009), Climate change: The El Niño with a difference, *Nature*, 461, 481–484, doi:10.1038/461481a
- **Bjerknes, J.** (1969). Atmospheric Teleconnections from the Equatorial Pacific. *Monthly Weather Review*, vol. 97, Na 3, pp: 163-172.
- **Capotondi, A., and Coauthors.** (2015). Understanding ENSO diversity, *Bull. Amer. Meteor. Soc.*; doi:10.1175/ BAMS-D-13-00115.1.
- **Chang P, Zhang L, Saravanan R, Vimont DJ, Chiang JCH, Ji L, Seidel H, Tippett MK** (2007). Pacific meridional mode and El Niño-Southern Oscillation. *Geophys Res Lett* 34:L16608. doi: [10.1029/2007GL030302](https://doi.org/10.1029/2007GL030302)
- **Chapin, F., Chapin, M., Matson, P., & Vitousek, P.** (2011). Principles of terrestrial ecosystem ecology (2nd ed.). New York: Springer.
- **Davis, R. E.** (1976). Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.*, 6, 249.
- **Dewitte B., Vazquez-Cuervo J., Goubanova K., Illig S., Takahashi K., Cambon G., Purca S., Correa D., Gutierrez D., Sifeddine A. and Ortlieb L.** (2012). Change in El Niño favours over 1958-2008: Implications for the long-term trend of the upwelling of Peru. *Deep Sea Research II*, doi:10.1016/j.dsr2.2012.04.011.
- **Dommenget, D., Bayr T., and Frauen C.** (2013). Analysis of the non-linearity in the pattern and time evolution of El Niño Southern Oscillation. *Climate Dyn.*, 40, 2825–2847, doi: 10.1007/s00382-012-1475-0.
- **Figueroa, N.** (2001). Física Básica y Medio Ambiente. Editorial Coveñas. Lima.
- **Gabler, R., Peterson J., and Trapasso L.** (2007). Essentials of Physical Geography, 8th Ed. Thomson Brooks/Cole. Belmont.
- **Gill, A. E.** (1982). Atmosphere–Ocean Dynamics. London: Academic Press.
- **Hoerling, M. P., Kumar A., and Zhong M.** (1997). El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate*, 10, 1769–1786
- **Horel, J.D.** (1984). Complex principal component analysis: theory and examples. *J. Clim. Appl. Meteorol.* 23,1660-1673.
- **Hotelling, H.** (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**:417–41.
- **Hsieh, W.** (2004). Nonlinear multivariate and time series analysis by neural network methods. *Rev. Geophys.*, 42, RG1003, doi:10.1029/2002RG000112.
- **Hsieh, W.** (2009). Machine Learning Methods in the Environmental Sciences. Cambridge Press.
- **Husson, F., Lê, S., Pagès, J.** (2017). *Exploratory Multivariate Analysis by Example Using R*. 2nd edition. Chapman & Hall/CRC.
- **Isaza, F. y Campos, D.** (2007). Cambio climático: glaciaciones y calentamiento global. Fundación Universidad de Bogotá Jorge Tadeo Lozano. Colombia.
- **Jolliffe, I. T.** (2002). Principal Component Analysis. New York: Springer.

- **Kao H., Yu J.** (2009). Contrasting Eastern Pacific and Central Pacific types of ENSO. *J. Clim.* 22:615–632
- **Katsaros, K. B.** (1980). The aqueous thermal boundary layer. *Boundary-Layer Meteorol.*, **18**, 107–127.
- **Kelly, K.** (1988). Comment on ‘Empirical orthogonal function analysis of advanced very high resolution radiometer surface temperature patterns in Santa Barbara Channel’ by G.S.E. Lagerloef and R.L. Bernstein. *Journal of Geophysical Research*, **93**(C12):15743–54.
- **Kidson, J.W.** (1975). Eigenvector Analysis of Monthly Mean Surface Data. *Monthly Weather Review*, 103(3):177-186.
- **Kim, H., Webster, P., and Curry, J.** (2009), Impact of shifting patterns of Pacific Ocean warming on North Atlantic tropical cyclones, *Science*, 325,77–80, doi:10.1126/science.1174062
- **Kug, J., Jin F., and An S.** (2009). Two types of El Niño events: Cold tongue El Niño and warm pool El Niño, *J. Clim.*, 22, 1499–1515.
- **Lagerloef, G., and Bernstein R.** (1988). Empirical orthogonal function analysis of advance very high resolution radiometer surface temperature patterns in Santa Barbara channel. *J. Geophys. Res.*, 93, 6863–6873.
- **Lee, T., and McPhaden, M.** (2010). Increasing intensity of El Niño in the central-equatorial Pacific, *Geophys. Res. Lett.*, 37, L14603, doi:10.1029/2010GL044007.
- **Lorenz, E.** (1956). Empirical orthogonal functions and statistical weather prediction. Sci. rep. no. 1, Statistical Forecasting Project, Dept. of Meteorology, MIT.
- **Monahan, A.** (2001). Nonlinear principal component analysis: Tropical Indo–Pacific Sea surface temperature and sea level pressure. *J. Climate*, 14, 219–233.
- **Monahan, A., and Dai, A.** (2004). The spatial and temporal structure of ENSO nonlinearity. *J. Climate*, 17, 3026–3036.
- **Murray, C., Mueller, L., and Zwally, H.** (1984). Matrix partitioning and EOF/principal component analysis of Antarctic sea ice brightness temperatures. National Aeronautics and Space Administration, Goddard Space Flight Center.
- **Neelin, J. D. & Latif, M.** (1998). El Niño Dynamics. *Physics Today*, 32-36.
- **Neumann, G., Pierson, W.** (1966). *Principles of physical oceanography*. Englewood Cliffs, NJ: Prentice-Hall.
- **North G. R., Bell, T. L., Cahalan, R. F. and Moeng, F. J.** (1982). Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110:699–706.
- **North, G. R.** (1984). Empirical orthogonal functions and normal modes. *J. Atmos. Sci.*, 41, 879–887.
- **Overland, J. E., and Preisendorfer R.** (1981). A significance test for principal components applied to a cyclone climatology. *Mon. Wea. Rev.*, 110, 1–4.
- **Pearson, K.** (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Ser. 6, **2**:559–72

- **Peixoto, J.P., and Oort A.H.** (1992). *Physics of Climate*, New York. American Inst. Physics, 520 pp.
- **Philander, S. G.** (1990). *El Niño, La Niña, and the Southern Oscillation*. San Diego: Academic Press.
- **Preisendorfer, R. and Mobley, C.** (1988). *Principal component analysis in meteorology and oceanography*. Elsevier Amsterdam.
- **Rasmusson, E. M., and Carpenter, T. H.** (1982). Variation in tropical sea surface temperature and surface wind fields associated with Southern Oscillation/El Niño, *Mon. Wea. Rev.*, *110*, 354-384.
- **Richman M.B.** (1986). Rotation of Principal Components, *Journal of Climatology*, *6*, 293–335.
- **Rohli, R. and Vega, A.** (2018). *Climatology* (4th. ed.), Jones & Bartlett.
- **Seelye M.** (2014). *An Introduction to Ocean Remote Sensing* University of Washington. Second edition. Cambridge, ISBN:9780521802802.
- **Shlens, A.** (2005). Tutorial on Principal Component Analysis. Copy retrieved from: <http://www.cs.cmu.edu/~elaw/papers/pca.pdf>
- **Strahler, A.** (2010) *Introducing Physical Geography*, 5th Edition. Wiley.
- **Strang, G.** (2005). *Linear Algebra and Its Applications*. Pacific Grove, CA: Brooks Cole.
- **Takahashi, K., Montecinos, A., Goubanova, K. and Dewitte, B.** (2011). ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophys. Res. Lett.*, *38*, L10704, doi:10.1029/2011GL047364.
- **Trenberth, K.** (1975). A quasi-biennial standing wave in the southern hemisphere and interrelations with sea surface temperature. *Quart. J. Roy. Meteor. Soc.* *101*, 55-74.
- **Venegas, S.** (2001) Statistical methods for signal detection in climate. DCESS Report No 2.
- **Vimont, D., Wallace, J., Battisti, D.** (2003). The seasonal footprinting mechanism in the Pacific: implications for ENSO. *J Clim* *16*(16):2668–2675
- **Wallace, J.M., Dickinson, R.E.** (1972). Empirical orthogonal representation of time series in the frequency domain. Part I: theoretical considerations. *J. Appl. Meteor.* *11*, 887-892.
- **Walton, C. C., Pichel, W. G., & Sapper, J. F.** (1998). The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *J. Geophys. Res.*, *103*, 27999–28012.
- **Weare, B., Navato, A. and Newell, R.** (1976). Empirical Orthogonal Analysis of Pacific Sea Surface Temperatures, *Journal of Physical Oceanography*: Vol. 6, No. 5, pp. 671–678.
- **Weare, B.C., Nasstrom, J.S.** (1982). Examples of extended empirical orthogonal function analyses. *Mon. Weath. Rev.* *110*, 481-485.

- **Weng, H., Behera, S. and Yamagata, T.** (2009). Anomalous winter climate conditions in the Pacific Rim during recent El Niño Modoki and El Niño events, *Clim. Dyn.*, 32, 663–674, doi:10.1007/s00382-0080394-6.
- **Wilks, D.** (1995). *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- **Yeh, S., Kug, J., Dewitte, B., Kwon, M., Kirtman, B., and Jin, F.** (2009). El Niño in a changing climate, *Nature*, 461, 511–515.
- **Yu, J., and Kim, S.** (2010). Three evolution, patterns of central Pacific El Niño, *Geophys. Res. Lett.*, 37, L08706, doi:10.1029/2010GL042810.